

From: Portner, Ginny
Sent: Friday, October 04, 2002 11:10 AM
To: STIC-Biotech/ChemLib
Subject: 09/848909

Please search SEQ ID No 21 and homologs thereto. (protein claims) thanks,

Ginny Portner
CM1, Art Unit 1645
Room 7e13
Mail box 7e12
(703) 308-7543

Interference done as
well 10/02
vso

Edward Hart
Technical Info. Specialist
STIC/Biotech
CMI 6B02 Tel: 305-9203

6329186
SEQ 8

5591631 SEQ 12
5677274 SEQ 12
5677274 SEQ 31

PCT/US 9401624
SEQ 12

acc #
AAR 60193
AAR 60183

WO 9418332
1994

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 10/4/02
Date Completed: 10/4/02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: 1
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: PR
WWW/Internet: _____
Other (specify): _____

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: October 4, 2002, 14:34:15 ; Search time 34.92 Seconds
(without alignments)
2341.074 Million cell updates/sec

Title: US-09-848-909-21

Perfect score: 3778
Sequence: 1 EVKQENRLMSESSSQGL.....STNGIKTLIPSKKGEIGZ 736

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: A_Geneseq_032802.*
2: /SIDSI/gcgdata/hold-geneseq/geneexp-emb1/AA1980.DAT.*
3: /SIDSI/gcgdata/hold-geneexp-emb1/AA1981.DAT.*
4: /SIDSI/gcgdata/hold-geneexp-emb1/AA1982.DAT.*
5: /SIDSI/gcgdata/hold-geneexp-emb1/AA1983.DAT.*
6: /SIDSI/gcgdata/hold-geneexp-emb1/AA1984.DAT.*
7: /SIDSI/gcgdata/hold-geneexp-emb1/AA1985.DAT.*
8: /SIDSI/gcgdata/hold-geneexp-emb1/AA1986.DAT.*
9: /SIDSI/gcgdata/hold-geneexp-emb1/AA1987.DAT.*
10: /SIDSI/gcgdata/hold-geneexp-emb1/AA1988.DAT.*
11: /SIDSI/gcgdata/hold-geneexp-emb1/AA1989.DAT.*
12: /SIDSI/gcgdata/hold-geneexp-emb1/AA1990.DAT.*
13: /SIDSI/gcgdata/hold-geneexp-emb1/AA1991.DAT.*
14: /SIDSI/gcgdata/hold-geneexp-emb1/AA1992.DAT.*
15: /SIDSI/gcgdata/hold-geneexp-emb1/AA1993.DAT.*
16: /SIDSI/gcgdata/hold-geneexp-emb1/AA1994.DAT.*
17: /SIDSI/gcgdata/hold-geneexp-emb1/AA1995.DAT.*
18: /SIDSI/gcgdata/hold-geneexp-emb1/AA1996.DAT.*
19: /SIDSI/gcgdata/hold-geneexp-emb1/AA1997.DAT.*
20: /SIDSI/gcgdata/hold-geneexp-emb1/AA1998.DAT.*
21: /SIDSI/gcgdata/hold-geneexp-emb1/AA1999.DAT.*
22: /SIDSI/gcgdata/hold-geneexp-emb1/AA2000.DAT.*
23: /SIDSI/gcgdata/hold-geneexp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3774	99.9	735	15	AA60179
2	3774	99.9	736	21	AA556959
3	3774	99.9	763	21	AA556960
4	3774	99.9	764	21	AA556961
5	3774	99.9	764	22	AA87306
6	3729.5	98.7	903	15	AA60193
7	3597	95.2	719	15	AA60183
8	2919	77.3	569	21	AA556961
9	855.5	22.6	1052	22	AA607903
10	854	22.6	1032	22	AA607901
11	850	22.5	1092	22	AA607900

12	841.5	22.3	1112	22	AA607902	C. botulinum C2 tr
13	792.5	21.0	880	19	AA60224	Bacillus thuringie
14	792.5	21.0	881	21	AA59277	MIS toxin from B.
15	790	20.9	884	17	AA591239	B. cereus VIP1A(a)
16	790	20.9	884	18	AA591509	B. cereus VIP1A(a)
17	790	20.9	884	19	AA591512	100 kDa VIP1A(a) t
18	790	20.9	1346	17	AA591245	VIP2A(a) and VIP1A
19	790	20.9	1346	18	AA591513	B. cereus VIP1A(a)
20	790	20.9	1346	19	AA591515	VIP1A(a)/VIP2A(a)
21	789	20.9	852	17	AA591246	VIP1A(a) protein
22	789	20.9	852	18	AA591516	Maize optimised B.
23	789	20.9	852	19	AA591517	VIP2A(a)-VIP1A(a)
24	789	20.9	1338	17	AA591247	VIP2A(a)-VIP1A(a)
25	789	20.9	1338	18	AA591520	Maize optimised B.
26	789	20.9	1338	19	AA591521	VIP2A(a)/VIP1A(a)
27	785	20.8	884	15	AA591793	Bacillus cereus 10
28	775.5	20.5	784	21	AA592948	MIS-8 toxin from B
29	766.5	20.3	860	21	AA59282	B. thuringiensis V
30	756.5	20.0	834	17	AA591242	B. thuringiensis V
31	756.5	20.0	834	18	AA591512	Vegetative insect
32	756.5	20.0	834	19	AA591512	Bacillus thuringie
33	608.5	16.1	425	19	AA591240	B. cereus thuringie
34	603	16.0	667	17	AA591240	B. cereus 80 kD VI
35	603	16.0	667	18	AA591510	80 kDa VIP1A(a) to
36	603	16.0	667	19	AA591510	Bacillus thuringie
37	601	15.9	667	15	AA591794	Bacillus thuringie
38	588	15.6	357	19	AA591240	Bacillus thuringie
39	581	15.4	357	19	AA591240	Bacillus thuringie
40	572	15.1	357	19	AA591240	Bacillus thuringie
41	567	15.0	357	19	AA591240	Bacillus thuringie
42	553	14.6	357	19	AA591240	Bacillus thuringie
43	471	12.5	327	19	AA591240	Bacillus thuringie
44	468	12.4	327	19	AA591240	Bacillus thuringie
45	435.5	11.5	348	19	AA591240	Bacillus thuringie

ALIGNMENTS

RESULT 1	AA60179	standard; Protein: 735 AA.
ID	AA60179	
AC	AA60179	
XX	03-APR-1995	(first entry)
DE	Protective antigen of Bacillus anthracis.	
XX		
KW	Anthrax; Bacillus anthracis; fusion protein; protective antigen; protective antigen; cell killing; targeting; pathogen; intracellular; HIV; human immunodeficiency virus; toxin.	
XX		
OS	Bacillus anthracis.	
XX		
PN	W09418332-A.	
XX		
PD	18-AUG-1994.	
XX		
PF	14-FEB-1994; 94MO-US01624.	
XX		
PR	12-FEB-1993; 93US-0021601.	
XX	25-JUN-1993; 93US-0082849.	
XX	(USHS) US DEPT HEALTH & HUMAN SERVICES.	
XX		
PI	Aroza N, Kilmpel K, Leppla SH, Nichols PJ, Smith Y;	
XX	WPI; 1994-279753/34.	
DR	N-PSDB; AA070180.	
XX		
PT	Nucleic acid encoding anthrax toxin fusion protein - useful for targeting toxin to specific cells, eg for killing tumour cells	

PT or HIV-infected cells

PS Disclosure; Page 81-83; 12app; English.

CC The sequence encoding the protective antigen of *Bacillus anthracis*
CC may be used in the construction of a nucleic acid which encodes a
CC fusion protein comprising the anthrax protective antigen binding
CC domain of the native anthrax lethal factor and a sequence encoding an
CC activity inducing domain of a second protein. The fusion proteins
CC are useful for the specific killing of tumour cells or the killing of
CC cells infected with intracellular pathogens, especially HIV.

XX Sequence 735 AA:

Query Match 99.9%; Score 3774; DB 15; Length 735;
Best Local Similarity 100.0%; Pred. No. 4.2e-245;
Matches 735; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVKQENRLNSESSESSOGILGYFSDLNFOAPMVVYSSSTTGDLSPSELENIPSENOYF 60
DB 1 evkqenrlnesesssgllgyfslndfagpmvvtsttdglstpselenipengyf 60
QY 61 OSATMSGFIKKKSDXYTFATSDNHYTMVDDQEVYINKASNSKRIKELKGLYQIKIY 120
DB 61 gsatmsgfiikkdsdeytfatsadnhvtdqevyinkasnskrlkelyqlyqiky 120
QY 121 QRENTEKGLDFKLYTDSQNKKEVYSSDNLOPELKOKSSNRKRSISAGTYPPDRN 180
DB 121 qrenptekglfdklytwsqnkkevysdnlolpelkoksnsnrkrsisagtyppdrn 180
QY 121 qrenptekglfdklytwsqnkkevysdnlolpelkoksnsnrkrsisagtyppdrn 180
QY 181 DGIPDSLEVEGYTVDVKNKRTFLSPWISNHEKGLJTKYKSSPEKWSADSPDEKVT 240
DB 181 dgipdsleveytvdvknkrtflspwlsnhkgljtkykspekwsadspdekevlt 240
QY 241 GRIDKNVSPKARHPLVAAPYIVHDMENIILSKNEQSTQNTDSEFTTISKMTSRTHT 300
DB 241 gridknvspearhplvaapyivhdmenniilskneqstqntdsefttiskmtsrtht 300
QY 301 SEVHGNNAEHAASFPIGVSAGFSNNSSTVAIDHSLSIAGERTMAETGTLNTADTARL 360
DB 301 sevhnnaehaasfpidgvsagfnsnstvalidhslsiagertmaetgltntadtarl 360
QY 361 NANIRRYNVTAPYVNLPTTSVYGNKQTLATIKAKENQSLQILANNNYPSKMLAPIA 420
DB 361 nanirrynvgtapyvnlpttsvlygnkqtlatikakengsqllapnnypskmlapia 420
QY 421 LNAQDPFSTPTNNYNQPLELEKTKOLRDPVOYGNATYNEENRANVDTGSNMEV 480
DB 421 lnaqdpfstptnnynqplelektrkolarpvoynatynneenranvdtgsnmev 480
QY 481 LPQIETTRKRIIFNGKOLNVERRIAANVPSDPLETTFKDMTLKALKAFGEPEPNCNL 540
DB 481 lpiqiettrkriifngkoldnverriaanvpsdplettfkdmtlkalkafgepncnl 540
QY 541 QYQKDIETFEFNDQOTSONIKNOALAEANATYIVLDKIKLAKANNILIRDKRFHYDR 600
DB 541 qyqkdietfeefndqotsoniknolaelaenatyivldkiklakannilirdkrfhydr 600
QY 601 NNIAVGADESVVKAHREVINNSTEGILNIDNDIKILSGIYIETDEGKLEKYEINRY 660
DB 601 nniavgadesvvkaahrevinsstegilnidndikilsgiyiavedteglekveinyry 660
QY 661 DMLNLSLRODKTFIDFKYNDKLPITISNPNYKYNVYAATKENTLINSNGDTSTNG 720
DB 661 dmlnlsldoktfidfkynyndklpitisnpykynvyaatkentlinsngdtstng 720
QY 721 IKRLIFSRRKGEIG 735
DB 721 ikrlifskkygeig 735

RESULT 2

AAV56959
ID AAV56959 standard; protein; 736 AA.

AC AAV56959;

XX 25-APR-2000 (first entry)

DE B. anthracis MAT-PA protein.

XX Bacillus anthracis protein; protective antigen; PA; MAT-PA; TPA-PA;
XX tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.

OS Bacillus anthracis.

PN WO200002522-A2.

PD 20-JAN-2000.

XX 09-JUL-1999; 99MO-US15568.

XX 10-JUL-1998; 98US-0092416.

PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.

XX Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;

XX WPI; 2000-182165/16.

DR N-P-SDB; AAZ56875.

PT Recombinant DNA construct useful as vaccines for anthrax, in producing
XX host cells for analyzing the drugs and agents inhibiting anthrax

XX Disclosure; Page 34; 35pp; English.

CC The invention provides a recombinant DNA construct that comprises a
CC vector and at least one nucleic acid (or its fragment) encoding a
CC combination of *Bacillus anthracis* proteins; selected from protective
CC antigen (PA), MAT-PA (PA with its secretory signals removed), TPA-PA
CC (PA with its secretory signals replaced with those of tissue plasminogen
CC activator) and PA63 (63 kDa PA). The construct is useful as a DNA vaccine
CC for anthrax and in producing infectious alpha virus particles. These
CC particles, expressing the B. anthracis proteins are useful also as
CC vaccines for anthrax. Host cells transformed with the construct are
CC useful for analyzing the effectiveness of drugs and agents that inhibit
CC anthrax or B. anthracis proteins. The present sequence represents a
CC B. anthracis MAT-PA protein.

XX Sequence 736 AA;

Query Match 99.9%; Score 3774; DB 21; Length 736;
Best Local Similarity 100.0%; Pred. No. 4.2e-245;
Matches 735; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EVKQENRLNSESSESSOGILGYFSDLNFOAPMVVYSSSTTGDLSPSELENIPSENOYF 60
DB 1 evkqenrlnesesssgllgyfslndfagpmvvtsttdglstpselenipengyf 61
QY 61 OSATMSGFIKKKSDXYTFATSDNHYTMVDDQEVYINKASNSKRIKELKGLYQIKIY 120
DB 61 gsatmsgfiikkdsdeytfatsadnhvtdqevyinkasnskrlkelyqlyqiky 121
QY 121 QRENTEKGLDFKLYTDSQNKKEVYSSDNLOPELKOKSSNRKRSISAGTYPPDRN 180
DB 121 qrenptekglfdklytwsqnkkevysdnlolpelkoksnsnrkrsisagtyppdrn 181
QY 181 DGIPDSLEVEGYTVDVKNKRTFLSPWISNHEKGLJTKYKSSPEKWSADSPDEKVT 240
DB 181 dgipdsleveytvdvknkrtflspwlsnhkgljtkykspekwsadspdekevlt 241
QY 241 GRIDKNVSPKARHPLVAAPYIVHDMENIILSKNEQSTQNTDSEFTTISKMTSRTHT 300
DB 241 gridknvspearhplvaapyivhdmenniilskneqstqntdsefttiskmtsrtht 301

```

OY 301 SEVHGAHVASFPGDGSVSGFSNSNSTVAIDHSLSLAGEFTVAEMGTADTARL 360
    |||||||
DB 302 sevbgnaevhasffidgysvsgfsnsstvaldhsislageretwaelmgintadtarl 361
OY 361 NANIRYVNTGAPLYNVLPFTSLVIGKNOTLATIRAKENOSQILAPNNYPSKMLAPIA 420
    |||||||
DB 362 naniryvntgcaplynvlpftslvlgknotlatikakenqslqllapnypskmlapia 421
OY 421 LNAODFSSPTITMNYNOFLEKTKQLRDTQVYGNATYFNGRVRVDTSNSSEV 480
    |||||||
DB 422 lnaodfssptitumnyngflekktqrlidatqygnatynengrvrvdtsnsew 481
OY 481 LPDIQETTARIIFNGKDLNVERRIAANPSDLETPEDMTLEAKTIAFENPNGL 540
    |||||||
DB 482 lpdigettarilifngkdlhvertiaavnsdpdletpdmulkealkiafienpognl 541
OY 541 QYQGRDITEFDFNFDOOTSNIKNOELANATYIVLDKIKLAKKNLIDRKFHYDR 600
    |||||||
DB 542 qyqgkditefndfngkdlhvertiaavnsdpdletpdmulkealkiafienpognl 601
OY 601 NNIAGADESVYKAEHREYVINSSTEGLLNIDKRIKLSGYIELEDEGLKEVINDRY 660
    |||||||
DB 602 nniavgadevsvkaehevinsstegllnidkdkllsgyiledeeglkevindry 661
OY 661 DMNLSLRDQKGTIDFKKYNDKPLIYISNPYKVNVAATKENTLINPSENGDTSTNG 720
    |||||||
DB 662 dmnlslsrqgktdfkfkyndkpliyisnpynvayavtkentlinpsengdstng 721
OY 721 IKKILIFSKGYEIG 735
    |||||||
DB 722 ikkiliifskgyeig 736

RESULT 3
AAV56960 standard; protein: 763 AA.
XX
AC AAV56960;
XX
DT 25-APR-2000 (first entry)
XX
DE B. anthracis TPA-PA protein.
XX
KV Bacillus anthracis protein; protective antigen: PA; MAT-PA; TPA-PA;
KW tissue plasminogen activator; PA63; vaccine; anthrax; antibacterial.
XX
OS Bacillus anthracis.
XX
PM W0200002522-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-0515568.
XX
PR 10-JUL-1998; 98US-0092416.
XX
PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;
XX
PI WPI; 2000-182165/16.
XX
DR N-PSDB; AAV56876.
XX
PT Recombinant DNA construct useful as vaccines for anthrax. In producing
XX host cells for analyzing the drugs and agents inhibiting anthrax
XX
PS Disclosure; Page 32; 35pp; English.
XX
CC The invention provides a recombinant DNA construct that comprises a
CC vector and at least one nucleic acid (or its fragment) encoding a
CC combination of Bacillus anthracis proteins, selected from protective
CC antigen (PA), MAT-PA (PA with its secretory signals removed), TPA-PA

```

```

CC (PA with its secretory signals replaced with those of tissue plasminogen
CC activator) and PA63 (63 kDa PA). The construct is useful as a DNA vaccine
CC for anthrax and in producing infectious alpha virus particles. These
CC particles, expressing the B. anthracis proteins are useful also as
CC vaccines for anthrax. Host cells transformed with the construct are
CC useful for analyzing the effectiveness of drugs and agents that inhibit
CC anthrax or B. anthracis proteins. The present sequence represents a
CC B. anthracis TPA-PA protein.
XX
XX Sequence 763 AA:
XX
OY 1 EYKQENRLNRSSESSOGILCYPSDNLROAPMYVTSSTGDLSTSSSLRNISENGYF 60
    |||||||
DB 29 eykqenrlnrsessogilgyrpsdnlrqapmvytsstgdlstssslrniseqyft 88
OY 61 QSALNSGFTKKKSDERYTATSDNHYTMVDDQEVYINKASNSNKRLKGRGLYQIKIY 120
    |||||||
DB 89 qsalnsqftkkksderytatdsadnhvumvddqevynkasnsnkrlkgrlyqikly 148
OY 121 QRENPTKGLDFKLYWTDSONKREYISSDNLQLEPKOKSSNSKRRSTSAQPIVPRDN 180
    |||||||
DB 149 qrenptekgldfklywtdsqnkreyissdnlqlpepkokssnskrrstsaqpivrdrn 208
OY 181 DGIPDLEEGYTVYKNNRFLSPWISNIHERKGLTRYKSSPEKSTASDPSDEKVT 240
    |||||||
DB 209 dgipdsleegytyvdknnrflspwlsnihekkglytkyspewstasdpdsekv 268
OY 241 GRIDKNVSPFARHPLVAAPYIVHDMENILSKNEQDONTDSEPTSKTSRTHT 300
    |||||||
DB 269 gridknvspfarhplvaapyivhdmennilskneqdotdseptsktsktsrtht 328
OY 301 SEVHGAHVASFPGDGSVSGFSNSNSTVAIDHSLSLAGEFTVAEMGTADTARL 360
    |||||||
DB 329 sevbgnaevhasffidgysvsgfsnsstvaldhsislageretwaelmgintadtarl 388
OY 361 NANIRYVNTGAPLYNVLPFTSLVIGKNOTLATIRAKENOSQILAPNNYPSKMLAPIA 420
    |||||||
DB 389 naniryvntgcaplynvlpftslvlgknotlatikakenqslqllapnypskmlapia 448
OY 421 LNAODFSSPTITMNYNOFLEKTKQLRDTQVYGNATYFNGRVRVDTSNSSEV 480
    |||||||
DB 449 lnaodfssptitumnyngflekktqrlidatqygnatynengrvrvdtsnsew 508
OY 481 LPDIQETTARIIFNGKDLNVERRIAANPSDLETPEDMTLEAKTIAFENPNGL 540
    |||||||
DB 509 lpdigettarilifngkdlhvertiaavnsdpdletpdmulkealkiafienpognl 568
OY 541 QYQGRDITEFDFNFDOOTSNIKNOELANATYIVLDKIKLAKKNLIDRKFHYDR 600
    |||||||
DB 569 qyqgkditefndfngkdlhvertiaavnsdpdletpdmulkealkiafienpognl 628
OY 601 NNIAGADESVYKAEHREYVINSSTEGLLNIDKRIKLSGYIELEDEGLKEVINDRY 660
    |||||||
DB 629 nniavgadevsvkaehevinsstegllnidkdkllsgyiledeeglkevindry 688
OY 661 DMNLSLRDQKGTIDFKKYNDKPLIYISNPYKVNVAATKENTLINPSENGDTSTNG 720
    |||||||
DB 689 dmnlslsrqgktdfkfkyndkpliyisnpynvayavtkentlinpsengdstng 748
OY 721 IKKILIFSKGYEIG 735
    |||||||
DB 749 ikkiliifskgyeig 763

RESULT 4
AAV56958 standard; protein: 764 AA.
XX

```

AC	AAY56958;
XX	25-APR-2000 (first entry)
DT	B. anthracis protective antigen (PA) protein.
XX	
DE	Bacillus anthracis protein; protective antigen; PA; MAT-PA; TPA-PA;
KW	tissue plasminogen activator; PA65; vaccine; anthrax; antibacterial.
OS	Bacillus anthracis.
XX	
PN	MO200002522-A2.
PD	20-JAN-2000.
FE	09-JUL-1999; 99WO-US15568.
PR	10-JUL-1998; 98US-0092416.
XX	(USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
PA	Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;
PI	WPI: 2000-182165/76.
DR	N-Psdb; AAZ56874.
XX	
PT	Recombinant DNA construct useful as vaccines for anthrax, in producing
PT	host cells for analyzing the drugs and agents inhibiting anthrax -
PS	Disclosure; Page 33; 35pp; English.
CC	The invention provides a recombinant DNA construct that comprises a
CC	vector and at least one nucleic acid (or its fragment) encoding a
CC	combination of Bacillus anthracis proteins, selected from protective
CC	antigen (PA), MAT-PA (PA with its secretory signals removed), tPA-PA
CC	(PA with its secretory signals replaced with those of tissue plasminogen
CC	activator) and PA65 (63 kDa PA). The construct is useful as a DNA vaccine
CC	for anthrax and in producing infectious alpha virus particles. These
CC	particles, expressing the B. anthracis proteins are useful also as
CC	vaccines for anthrax. Host cells transformed with the construct are
CC	useful for analyzing the effectiveness of drugs and agents that inhibit
CC	B. anthracis PA protein.
XX	
Sequence	764 AA:
Query Match	99.9%; Score 3174; DB 21; Length 764;
Best Local Similarity	100.0%; Pred. No. 4,4e-245;
Matches 735; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 EYKQENRLNLESESSSGCLAGYVESDLNQAPMVVTSSTTGDSLSPSELENIPSENOYF 60
DB	30 evkgenrlleessessglllyyfsdlndfapmvvtsttgdsllspseleinspsengyf 89
OY	61 GSVAISGRIVAKKSDERYFAFSADNVHVTMMVDQDEVINKASNNKIRLEKGRLYQIKIY 120
DB	90 gsalwsgfflvkkksderylfatsadnhvctmvdqdevlnkasnnkirklegrylqikiy 149
OY	121 QRENTTEKGDFLKTWTDSOKKKRYTSSNQLDELTKOKSSNRKRKSTSGATPYDPBDN 180
DB	150 grempetkgdfdkltwtidsqnkkrveissdnqlpelkqssnrrkstsaagptypddcn 209
OY	181 DDPIPSLEVGYGVYVKKKRTFLSPGISIHHEKGLTKKYSSPEKMSTAADPSPERKYT 240
DB	210 ddippslevgygvayvkkrflfspalshhekgyltkkysspekmsaaadpysderkyt 269
OY	241 GTIDKNNSPFAKRPFLYAALPVHYWDENITLLSKNEDOSTONTDSTRTISKNTSTSRIHT 300
DB	270 gtidknnspearlpvlaayplivhdmenitllsknedgstncdstcrtlsknstsrht 329
OY	301 SEVHGNAEHAASFDDIGSVSAFGSNSSSTVAIDHSLSLAGERTWAETMGLTADPARL 360
DB	

Accession	Protein	Location/Qualifiers
D8330	sevginaevnasfidiqgsvsagfmsnosstvaldhs	lslagerlwaetnglntadearl 369
QY361	NANIRYVNGCTAPYINVLPTTSLVLRKNGKOTLATIKKENOL	SQI:APNNYPSKMLAPIA 420
D8390	nanlryngtclpnyvptstslvlsknglatlakenglsq	lappnyypskmlapia 449
QY421	LKADODSSSPITMANYNOFLTEKTKOLRLTDQVGNIAATYNS	NSRPVVDGSSMSSEV 480
D8450	lmgddsscpilmymyngflektrqlidtdqygnla	cytengrvrvdtsnasev 509
QY481	LPQIDETTAITIPKGNKDLNVERIAANVSDPELTAKPDMTK	:AKIAFGNEPKNOL 540
D8510	lpgjettatallngskdlnverriaavpsdpletcpdm	lksxlkafgnepgnl 569
QY541	OTGKDIITEFDEPNDOOTSONINQOLAEINMTNITYLEK	ILNAKKNMILIRDKRFHYDR 600
D8570	gyqgkditefdhfdqgtsqnlknqlaeinahtyvlck	lmaaknmllidckrfhydr 629
QY601	NNIAVGADESVYKEAHRVINSSTBGLNLINDXDIRKISG	YV:ILDEPGLKEVINYDR 660
D8630	nniaavgadesvyeahevinssteglllnidckrlfslsg	ylveidteglkevlndry 669
QY661	DMLNTSSLRDDGTFIDKRYKNKLEPLYSNPKNVYVAVKE	HLIINSEKDGSTNG 720
D8690	dmlntslrgdgtfidefkkyndkplysnpyknyvavke	hlipseogdstng 749
QY721	IKKILIRSKRGYEIG 735	
D8750	IKKILIRSKRGYEIG 764	
RESULT 5		
AD847306	standard: Protein: 764 AA.	
AD847306	standard: Protein: 764 AA.	
AA847306:		
D829-AUG-2001	(first entry)	
DE	wild type B. anthracis protective antigen.	
KW	lethal factor; LF: immunogen; LF4: protective antigen; PA: DNA vaccine	
KX	humoral; cell-mediated; immune memory response.	
OS	Bacillus anthracis.	
XX		
XX	key	Location/Qualifiers
XX	Peptide	1..29
XX		/label= Signal peptide
XX		/note= "Not given in the specification"
XX	Protein	30..764
XX	Peptide	/label= PA
XX		204..764
XX		/label= pCPA
PN	MO200145639-A2.	
XX	28-JUN-2001.	
XX	21-DEC-2000; 2000MO-US34912.	
XX	22-DEC-1999; 99US-011459.	
XX	(OHIS.) UNIV OHIO STATE RES FOUND.	
XX	(GALL.) GALLOWAY D R.	
XX	(MATEJ.) MATEJZON A J.	
XX	Galloway DR, Mateczun AJ.	
XX	WPI: 2001-408540/43.	
XX	N-PSDB: AAC86016.	
XX	Protecting animal against lethal infection with Bacillus anthracis, by	

PT administering wild type or mutated form of Bacillus anthracis lethal
PT factor protein or its fragment or a nucleic acid encoding the mutated
protein -

XX Claim 5; Fig 2; 33pp; English.

CC This sequence shows the B. anthracis protective antigen (PA). An
CC immunogenic fragment of PA, pcPA, can be used to produce an immune
CC response which protects an animal against lethal injection with
CC Bacillus anthracis. DNA encoding the B. anthracis PA can be used
CC in conjunction with DNA encoding the lethal factor (LF) in a
CC DNA vaccine. Using a DNA vaccine which encodes the mutated LF protein
CC or its fragment alone or in combination with a DNA encoding the PA protein
CC or its fragment, both components (humoral and cell-mediated) of the
CC immune system are stimulated, which results in longer term immune
CC memory response. The combined use of a mutated LF and PA gene or their
CC fragments results in a higher level of immune response, as judged by
CC overall serum antibody titers for LF and PA antigens, than the use of
CC either LF or PA genes in separate immunizations.

XX Sequence 764 AA:

Query Match 99.9%; Score 3774; DB 22; Length 764;
Best Local Similarity 100.0%; Pred. No. 4,4e-245;
Matches 735; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EVKQENRLNSESSESSQGLGYFFSDLNFOAPWVTSTGDLSPSSLENIPISENOYF 60
DB 30 EVKQENRLNSESSESSQGLGYFFSDLNFOAPWVTSTGDLSPSSLENIPISENOYF 89
OY 61 QSAIWSGFIKVKKSDYFATSNADNHTVMWDDQEVINKANSNKRILEKGRLYQIKQY 120
DB 90 QSAIWSGFIKVKKSDYFATSNADNHTVMWDDQEVINKANSNKRILEKGRLYQIKQY 149
OY 121 QRENPTKGLDFLYWTDSONKKEVYISSDNLOPELKOKSSNKRKSTSAEPYVDPDRN 180
DB 150 QRENPTKGLDFLYWTDSONKKEVYISSDNLOPELKOKSSNKRKSTSAEPYVDPDRN 209
OY 181 DGIPLDLEVEGYTDVKNKRTFLSPMISNTHKKGKTKYSKPKKSTASDPSPFEVY 240
DB 210 DGIPLDLEVEGYTDVKNKRTFLSPMISNTHKKGKTKYSKPKKSTASDPSPFEVY 269
OY 241 GRIDKRVSPKAPRLVNAAYIVHNDKEMILSKNEQSNQVTDETRTTSKRTSKRT 300
DB 270 GRIDKRVSPKAPRLVNAAYIVHNDKEMILSKNEQSNQVTDETRTTSKRTSKRT 329
OY 301 SEVHGNAEYHASFDDIGSVSGFSNSNSTVAIDHSLSLAGEETMAETMGINTADARL 360
DB 330 SEVHGNAEYHASFDDIGSVSGFSNSNSTVAIDHSLSLAGEETMAETMGINTADARL 389
OY 361 NANIRRYVNTGAPLYNVLPPTSLVGNKQTLATIKAKENQLSQIILAPNNYPSKMLAPIA 420
DB 390 NANIRRYVNTGAPLYNVLPPTSLVGNKQTLATIKAKENQLSQIILAPNNYPSKMLAPIA 449
OY 421 LNAQDDFSSTPTWNNYNOGLEKTKQLRLDQVGNIAVFNENGRVAVDTGSGMAGEV 480
DB 450 LNAQDDFSSTPTWNNYNOGLEKTKQLRLDQVGNIAVFNENGRVAVDTGSGMAGEV 509
OY 481 LPOIGETPARIIFPGKDLNVERRIAANESDPLETTRKPDTLKEAKTAFENPNPNL 540
DB 510 LPOIGETPARIIFPGKDLNVERRIAANESDPLETTRKPDTLKEAKTAFENPNPNL 569
OY 541 OYOGKRIEFPENPDQTSNKNKOLAFANATNIVYVDKTKIAAKNILLIDKREHFRDR 600
DB 570 OYOGKRIEFPENPDQTSNKNKOLAFANATNIVYVDKTKIAAKNILLIDKREHFRDR 629
OY 601 NRIAVAGDSYVKAHREYINSGGLLINDOKIRKILSGIYIEIDTGBGLKEYINDRY 660
DB 630 NRIAVAGDSYVKAHREYINSGGLLINDOKIRKILSGIYIEIDTGBGLKEYINDRY 689
OY 661 DMLNISLRODGKFTIDFKKYDKLPLIYISNPMYKXVNAVATKNTIINPSENGDTSTNG 720
DB 690 DMLNISLRODGKFTIDFKKYDKLPLIYISNPMYKXVNAVATKNTIINPSENGDTSTNG 749

DB 690 dmlnislrddgkftidfkkyndkpliyisnpykxvnavatkntiinpsengdstng 749
OY 721 IKKILIRSKKGYEIG 735
DB 750 IKKILIRSKKGYEIG 764

RESULT 6
AAR60183
ID AAR60183 standard; Protein: 903 AA.

XX AAR60183;

XX 04-APR-1995 (first entry)

XX PA(1-725)-----human CD4 fusion protein coding sequence.

XX Anthrax: Bacillus anthracis; fusion protein; lethal factor;

XX protective antigen; cell killing; targeting; pathogen;

XX Intracellular; HIV; human immunodeficiency virus; toxin;

XX Pseudomonas; exotoxin.

XX Bacillus anthracis.

XX Homo sapiens.

XX W09418332-A.

XX 18-AUG-1994.

XX 14-FEB-1994; 94MO-US01624.

XX 12-FEB-1993; 93US-0021601.

XX 25-JUN-1993; 93US-0082849.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Anora N, Klimpel K, Leppia SH, Nichols PJ, Slough Y;

XX WPI: 1994-279753/34.

XX N-PSDB; AAQ70184.

XX Nucleic acid encoding anthrax toxin fusion protein - useful for

XX targeting toxin to specific cells, eg for killing tumour cells

XX or HIV-infected cells

XX Disclosure: Page 100-103; 124pp; English.

XX This sequence is a fusion protein comprising amino acid residues

XX 1-725 of the anthrax protective antigen protein and residues 1-178

XX of human CD4, the portion which binds to gp120 on HIV infected

XX cells. Such fusion proteins may be useful for the specific killing

XX of tumour cells or the killing of cells infected with intracellular

XX pathogens, especially HIV, depending on their components.

XX Sequence 903 AA;

Query Match 98.7%; Score 3729.5; DB 15; Length 903;
Best Local Similarity 99.2%; Pred. No. 5.5e-242;
Matches 728; Conservative 1; Mismatches 2; Indels 3; Gaps 1;

OY 1 EVKQENRLNSESSESSQGLGYFFSDLNFOAPWVTSTGDLSPSSLENIPISENOYF 60
DB 1 EVKQENRLNSESSESSQGLGYFFSDLNFOAPWVTSTGDLSPSSLENIPISENOYF 60
OY 61 QSAIWSGFIKVKKSDYFATSNADNHTVMWDDQEVINKANSNKRILEKGRLYQIKQY 120
DB 61 QSAIWSGFIKVKKSDYFATSNADNHTVMWDDQEVINKANSNKRILEKGRLYQIKQY 120
OY 121 QRENPTKGLDFLYWTDSONKKEVYISSDNLOPELKOKSSNKRKSTSAEPYVDPDRN 180
DB 121 QRENPTKGLDFLYWTDSONKKEVYISSDNLOPELKOKSSNKRKSTSAEPYVDPDRN 180

```

QY 101 DGIIDSLFVETGYDVYKNNKRTFLSPWISNIEKKGLTKSSPEKNSADSPSYDEKYT 240
DB 181 dgiidslfvegydvynknkrtflspwisknlgklyksspeknsadspysdfxvt 240
QY 241 GRIDKNVSPARHPLVAAYPIVHVDMENTIISKNEOSTQNTDSEFTSKNTSTSTRT 300
DB 241 gridknvspearhplvaaypivhvdmentlisknedstqntdsetrlskntststrt 300
QY 301 SEVGNNAEVHASFPGIGSVAGFSNSNSTVAIDHSLSTAGEERTMAETGLTADTATL 360
DB 301 sevgnnaevhasfpgigsvagfnsnstvaidslslageretwaelgtatlatarl 360
QY 361 NANRYNVTGTPRIYVLPFTSLVIGKNQTLATIKAKENQLSQILAPNNYPSKNLAPIA 420
DB 361 nanrynvtgtpriyvlpftslvlgknqtlatikakenglsqilapnnypsknlapia 420
QY 421 LNAODDESSPTIMYNOFLEKTKQLRLDQVGNIAATYFNENGVRVYDGNMSWSEY 480
DB 421 lnaoddesptlmynoflektkqlrltdqvgniatynfengrvyvdgmswsev 480
QY 481 LPOIETARITENKDLNVERRIAANVPSDPLETTPKPMILAKIAFGNENGNL 540
DB 481 lpoietaaritengdlnverriaanvpsdplettpkpmilakiafgnenpnl 540
QY 541 OYOGKIDTFEPDNFOOTSONIKNOLAEINATNTIYVLDKIKLNKNNIILRDKRFHDR 600
DB 541 oyogkidtfepdnfootsoniknolaenatntiyvldkiklnknniilrdrkrfhdr 600
QY 601 NNIAVGADESVKEAHREVINNSTEGLINIDKDIRKILSGYVEIEDTEGLEVINDRY 660
DB 601 nniavgadesvkeahrevinnssteglinidkdirkilsgyveiedteglevindr 660
QY 661 DMLNSSLRODKTFIDFKYNDKLELYISNPKNYAVAVKENTILNPSNGDSTNG 720
DB 661 dmlnsslrodktfidfkyncklplyisnpknyavavkentilnpsngdstng 720
QY 721 IKKILU---IFSKKG 731
DB 721 ikkilkvvlygkkg 734

```

RESULT 7
AAR60193 standard: Protein; 719 AA.

AC AAR60193;
DT 04-APR-1995 (first entry)
XX Modified protective antigen of Bacillus anthracis.
DE Anthrax; Bacillus anthracis; fusion protein; lethal factor;
KM protective antigen; cell killing; targeting; pathogen;
KW intracellular; HIV; human immunodeficiency virus; toxin.
XX
OS Bacillus anthracis.
XX
PN MO9418332-A.
PD 18-AUG-1994.
XX
PE 14-FEB-1994; 94MO-USO1624.
PR 12-FEB-1993; 93US-0021601.
PR 25-JUN-1993; 93US-0082849.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Aroca N, Klimpel K, Leppla SH, Nichols PJ, Singh Y;
DR WPI: 1994-279753/34.
DR N-PSDB: AAQ70189.

PT Nucleic acid encoding anthrax toxin fusion protein - useful for
PT targeting toxin to specific cells, eg for killing tumour cells
PT or HIV-infected cells
XX
XX
PS Example 6; Page 114-115; 124pp; English.

The sequence encoding the protective antigen of Bacillus anthracis
CC may be used in the construction of a nucleic acid which encodes a
CC fusion protein comprising the anthrax protective antigen binding
CC domain of the native anthrax lethal factor and a sequence encoding an
CC activity inducing domain of a second protein. Such fusion proteins
CC are useful for the specific killing of tumour cells or the killing of
CC cells infected with intracellular pathogens, especially HIV.
CC depending on the second component. The protective antigen and other
CC toxins require proteolytic cleavage to acquire activity. Since some
CC cells infected with an intracellular pathogen possess an active
CC protease with quite a narrow substrate specificity e.g. HIV, the
CC intracellular pathogen specific protease site (See AAR60184-89). The
CC protease in cells that are infected with an intracellular pathogen
CC cleaves the modified toxin which is then rendered active and kills
CC the cell. This sequence is a modified Bacillus anthracis protective
CC antigen which has the amino acids originally at positions 102-171
CC replaced with the HIV protease cleavable sequence described in
CC AAR60186.
CC
XX
SQ Sequence 719 AA;

Query Match 95.2%; Score 3597; DB 15; Length 719;
Best Local Similarity 95.5%; Pred: No. 3-26-233;
Matches 706; Conservative 4; Mismatches 5; Indels 24; Gaps 2;

1 EVKQENRLNLSSESSSQGLGYFEDLNFOAPMVYSSSTGDLSTSSSELENIPSENOYE 60
DB 1 evkqenrlnlsesssgllyfyfedlnfapmvvrsstgdlstsselenipsengyf 60

61 QSAVNGCFKVKKSDERTFATSADNHVYMWVDDQEVINKASNSKIRLEKGLYQIKROY 120
DB 61 qsavngcfkvrksdertyfatsadnhvymvddqevinkasnskrllekgllyqikroy 120

121 QRENPEKGLDFKIXWTDSONKKEYISSDNLQPELKQKSSNS---RKRKSTSGPTVP 176
DB 121 qrenpekgldfkixwtdsonkkeyissdnlqlpekqkssns---rkrkstsgptvp 176

121 GRENPUEKYLIDKLYWCDsqnkkevissodnlqlpekqkssntcalimmgqgnfllgppvlp 180
DB 121 grenpuekyldklywcdsqnkkevissodnlqlpekqkssntcalimmgqgnfllgppvlp 180

177 DRDNDGIDPSLVEGTYVDVKNKRTFLSPWISNIEKKGLTKYKSSPEKNSADSPSYDF 236
DB 177 drdndgidpslvegytyvdvknkrtflspwisknlgklyksspeknsadspysdf 236

181 dindgidpslvegytyvdvknkrtflspwisknlgklyksspeknsadspysdf 240
DB 181 dindgidpslvegytyvdvknkrtflspwisknlgklyksspeknsadspysdf 240

237 EKYTGRIKNVSPARHPLVAAYPIVHVDMENTIISKNEOSTQNTDSEFTSKNTSTST 296
DB 237 ekvtgriknvsparhplvaaypivhvdmentlisknedstqntdsetrlskntstst 296

241 ekvtgridknvspearhplvaaypivhvdmentlisknedstqntdsetrlskntst 300
DB 241 ekvtgridknvspearhplvaaypivhvdmentlisknedstqntdsetrlskntst 300

297 RHTSEVGNNAEVHASFPGIGSVAGFSNSNSTVAIDHSLSTAGEERTMAETGLTAD 356
DB 297 rhtsevgnnaevaspfpgigsvagfnsnstvaidslslageretwaelgtatlat 356

301 rhtsevgnaevhasfpgigsvagfnsnstvaidslslageretwaelgtatlatad 360
DB 301 rhtsevgnaevhasfpgigsvagfnsnstvaidslslageretwaelgtatlatad 360

357 TARLNANRYNVTGTPRIYVLPFTSLVIGKNQTLATIKAKENQLSQILAPNNYPSKNL 416
DB 357 tarlnanrynvtgtpriyvlpftslvlgknqtlatikakenglsqilapnnypsknl 416

361 tarlnanrynvtgtpriyvlpftslvlgknqtlatikakenglsqilapnnypsknl 420
DB 361 tarlnanrynvtgtpriyvlpftslvlgknqtlatikakenglsqilapnnypsknl 420

417 APLAANADDESSPTIMYNOFLEKTKQLRLDQVGNIAATYFNENGVRVYDGNMSWSEY 476
DB 417 aplaanaddesptlmynoflektkqlrltdqvgniatynfengrvyvdgmswsev 476

421 aplaanaddesptlmynoflektkqlrltdqvgniatynfengrvyvdgmswsev 480
DB 421 aplaanaddesptlmynoflektkqlrltdqvgniatynfengrvyvdgmswsev 480

477 WSEVLPQIETARITENKDLNVERRIAANVPSDPLETTPKPMILAKIAFGNENGNL 536
DB 477 wsevlpoietaaritengdlnverriaanvpsdplettpkpmilakiafgnenpnl 536

461 wsevlpoietaaritengdlnverriaanvpsdplettpkpmilakiafgnenpnl 520
DB 461 wsevlpoietaaritengdlnverriaanvpsdplettpkpmilakiafgnenpnl 520

537 NGMLQYOGKIDTFEPDNFOOTSONIKNOLAEINATNTIYVLDKIKLNKNNIILRDKRF 596
DB 537 ngmlqyogkidtfepdnfootsoniknolaenatntiyvldkiklnknniilrdrkrf 596

521 ngmlqyogkidtfepdnfootsoniknolaenatntiyvldkiklnknniilrdrkrf 580
DB 521 ngmlqyogkidtfepdnfootsoniknolaenatntiyvldkiklnknniilrdrkrf 580

OY 597 HYDNNIAGADESVYKHAHEVINSSTEGILLNIDKDIRKILSGYIVETEDTEGLKEVI 656
 DB 581 hydnniavagadesvvykhahevinsstegillnldkdlrkilsgyiveledteglkevi 640
 OY 657 NDYVDMLNSSLRODGKTFIDFKRYNDKPLLYISNKNVFNVAVTEKNTIINPENGOT 716
 DB 641 ndyvdmlnsslrgdgkftfidefkryndkpllyisnknvfnvavteknentlinspengdt 700
 OY 717 STNGIKRILIFSRRGYEIG 735
 DB 701 stngikrllifsfkgyeig 719

RESULT 8

AAV56961
 ID AAV56961 standard; protein: 569 AA.

AC AAV56961:

DT 25-APR-2000 (first entry)
 DE B. anthracis PA63 protein.
 XX B. anthracis PA63 protein; protective antigen; PA; MAT-PA; TPA-PA;
 KM classe plasmidogen activator; PA63; vaccine; anthrax; antibacterial.
 XX Bacillus anthracis.
 OS WO200002522-A2.
 PN 20-JAN-2000.
 PD 09-JUL-1999; 99WO-US15568.
 PF 10-JUL-1998; 98US-0092416.
 PR (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
 XX Lee JS, Pushko P, Parker MD, Smith JF, Welkos SL;
 PI WPI: 2000-182165/16.
 DR N-PSDB; AAV56877.

PT Recombinant DNA construct useful as vaccines for anthrax, in producing
 host cells for analyzing the drugs and agents inhibiting anthrax
 XX Disclosure: Page 35; 35pp; English.

XX The invention provides a recombinant DNA construct that comprises a
 CC vector and at least one nucleic acid (or its fragment) encoding a
 CC combination of Bacillus anthracis proteins, selected from protective
 CC antigen (PA), MAT-PA (PA with its secretory signals removed), TPA-PA
 CC (PA with its secretory signals replaced with those of tissue plasminogen
 CC activator) and PA63 (63 kDa PA). The construct is useful as a DNA vaccine
 CC for anthrax and in producing infectious alpha virus particles. These
 CC particles, expressing the B. anthracis proteins are useful also as
 CC vaccines for anthrax. Host cells transformed with the construct are
 CC useful for analyzing the effectiveness of drugs and agents that inhibit
 CC anthrax or B. anthracis proteins. The present sequence represents a
 CC B. anthracis PA63 protein.

XX Sequence: 569 AA;

Query Match 77.3%; Score 2919; DB 21; Length 569;
 Best Local Similarity 100.0%; Pred No. 8, 2e-188;
 Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 168 STAGSPVVDNRDNCIDPDSLEVGCTVDYKRRKRTFLSPWISNHEKGLTKRYSSPEKMS 227
 DB 2 stagspvpvdndpdslevegtydvkrrkrtflspwlsnhkgltkrysspekms 61
 OY 228 TASDPYSDFEKYTKRDKNVSPEARHPLVAAYPIVHWDMENITLISKNDSTONTDSETR 287

DB 62 tasdpysdfekytkrdknvspearhplvaaypivhwdmenitlskndstqntdsestr 121
 OY 288 TISKNTSTSRHTSEVHNAEVAHASFEDIGSVAGSPNSNSVAADHSLSLAGEPTMA 347
 DB 122 tiskntstsrhtsevhnavehnaehasfedigsvagspnsnsvaadhsislagerpta 181
 OY 348 ETMGINTADTARLANIRRYVNTGTAPYVNLPTSLVGRKQIATLTKAKENOLSOILAP 407
 DB 182 etmgintadearlanirryvntgtapynlptslvgrkqiatltkakenolsqilap 241
 OY 408 NNYPSTKRLAPIALNODDPSPTTNVNOFLEKTKOLRLPDQVGNATYNPENG 467
 DB 242 nnypstkrllapialnoddpspttnvnoflektrkollrpdqvgnatynpeng 301
 OY 468 RRVNPTGSMSEVLRIOETTRAKIIFNGKDLNVEPRLAANPSELETPRDMTKLEAL 527
 DB 302 rrvnptgsmsevlrloettrakiifngkdlnveprrlaanpseletpdmtrkaleal 361
 OY 528 KTAGFNPENGNLQOGKDTFEPDNFOOTSONIKNOLAEALNNTIVVLDKIKLNAK 587
 DB 362 ktagfnpngnlqygkdtfepdnfootsniknolaelnntivvldkiklnakm 421
 OY 588 NILIRDKRPHYDRNNIAGADESVYKHAHEVINSSTEGILLNIDKDIRKILSGYIVE 647
 DB 422 nilirdkrphydrrnniavagadesvvykhahevinsstegillnldkdlrkilsgyive 481
 OY 648 DTEGLKEVINDRYDMLNSSLRODGKTFIDFKRYNDKPLLYISNKNVFNVAVTEKNTI 707
 DB 482 dteglkevindrlydmlnsslrgdgkftfidefkryndkpllyisnknvfnvavteknentl 541
 OY 708 INPENGDTSTNGIKRILIFSRRGYEIG 735
 DB 542 inspengdtstngikrllifsfkgyeig 569

RESULT 9

AAE07903
 ID AAE07903 standard; Protein: 1052 AA.

AC AAE07903;

DT 01-NOV-2001 (first entry)

XX C. botulinum C2 translocation domain with Tent binding domain #2.
 DE C. botulinum C2 translocation domain with Tent binding domain #2.
 KM Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
 KM tumour; infection; neurodegenerative disease; gene therapy;
 XX botulinum neurotoxin; tetanus neurotoxin; Tent.

XX Clostridium botulinum.
 OS Clostridium tetani.

PN WO200158936-A2.

XX 16-AUG-2001.

XX 04-DEC-2000; 2000WO-GB04644.

XX 02-DEC-1999; 99GB-0028530.

XX 07-APR-2000; 2000GB-0008658.

XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.

XX Shone CC, Sutton JM, Silman N;

XX WPI: 2001-514643/56.

PT New non toxic polypeptide for delivery of a therapeutic agent for the
 PT treatment of a CNS disorder comprising a binding domain that
 PT translocates the therapeutic agent into the neuronal cells -
 XX Example 2; page 50; 50pp; English.

XX The invention relates to a non toxic polypeptide, for delivery of a
 CC therapeutic agent to a neuronal cell, which comprises a binding domain
 CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
 CC as HC) that binds to the neuronal cell and a translocation domain (amino
 CC terminal half of HC, designated as HN), that translocates the therapeutic
 CC agent into the neuronal cell, where the translocation domain is not a HN
 CC domain of a clostridial neurotoxin and is not a fragment or derivative of
 CC a HN domain of a clostridial toxin. Polypeptides of the invention are
 CC useful for the treatment of a disease state associated with neuronal
 CC cells. The polypeptide constructs are useful for delivering therapeutic
 CC substances to neuronal cells. They are useful to treat disorders of the
 CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
 CC and infection. They are also useful in gene therapy. The present sequence
 CC is C. botulinum C2 enterotoxin translocation domain with tetanus
 CC neurotoxin (Tent) binding domain used in the exemplification of the
 CC invention.
 XX
 XX Sequence 1052 AA;

Query Match 22.6%; Score 855.5; DB 22; Length 1052;
 Best Local Similarity 33.1%; Pred. No. 9.5e-49;
 Matches 253; Conservative 138; Mismatches 297; Indels 77; Gaps 29;

QY 3 KOENRLNSES--SSQGLGYFSDLPQAPWYSSSTGDSIPSELENIPEENQYF 60
 DB 4 klenysvknknyftinglmgyfien-dfnlnlspcltgnlftskedinslg-nkll 61
 QY 61 QSATMSGPIKVKKSDERTFATSADN-HYTMWVDQEVYKASN-SNKIRLEKGLYQIKI 118
 DB 62 ksarwagllkpslftgyelstnspcrvel--ngelfnlstntsvnlhlgnydydlrl 118
 QY 119 -QYORENPTKGLD-FKIYWTDSQNKKEVSSDNLQPELKOKSSNRKRRTSAGPTV- 175
 DB 119 eqlmseengllknyegllkylwetadllkelpsevl-----lkpynsnekskfpinlfl 174
 QY 176 -----PRONDGIPDLEVEGTYVDVKNKRFPLSPMTSNHKKKLLPYKSSPEK 225
 DB 175 snaklkanaarldtdgldpewelingytwmngkavawdkfaan-----gykkysnpgfk 229
 QY 226 WSTASPDYSPDFKVTGKIDKNSPEARHPLVAAPVYHMDNIIISKNDGQNTQNTSE 285
 DB 230 pctandpydfdkvsgqlpsvswaiaidpmisayplvygmerlvakse-----ltlids 285
 QY 286 TRTISKNTSTSTHSEVHGNAPYHASFQDGC-----SYSGFENSNSSTVAIDSL 338
 DB 286 tkmsksthstntnly--gaevsgslqlagglfpyfmsasanyshcwnqstcvdcl- 342
 QY 339 SLAGETAETAGTADTARLNAIRVYNGTAPVYVNLPTTSVLGKQVLATITAKE 398
 DB 343 --tge-stfgslstnhtgasylnphlryngtapyvncptcltvdck-qvavatkqge 398
 QY 399 NQLSQILAPNNYPSKMLADIALNMODFSSPTPTMNYNQLEKTRQLRLDTPQVYGN 458
 DB 399 sllydyllngpylypllgpemahtndfssrllpnyqlksldngtvmststsgfyn 458
 QY 459 IATYFNGRVRHVDGNSMSEVLPOIOTETARI--FNKDLNIVERRIAAVNPSPLET 516
 DB 459 fakyn-sngnlyvtd-gnwgylygtlkstlaltlfsfgqtgya---vvapufedpedk 513
 QY 517 TKPDMTLKALKIAFGFNEPNCNLOQYOKDIT--EFDNPDQOTSONIKNOIAELNATN 573
 DB 514 t-pltlegalkaleknkgfyfngljstskneklgyfldstntndfengqlktackd 572
 QY 574 IYVYLKIKLANKKNIILIRDKRPHYDRNNINAVGADSVYKAEHREHVNSTGELLNDK 633
 DB 573 lmbcl--lkrn--mlllvkyltkenislnkldcwyneddivlklkstlndldnn 628
 QY 634 DIRRIISGV---IYIEDTEGKLEVINDRYMDNIISSLSQOKQFTIDFKYNDKPLPLYS 690
 DB 639 dltsdtslsgnssvltypdaqlvpglngaahlnyvnesevrvhkanmli-eyndmli- 692

QY 691 NPNKYNNYATKENTINPSENGDTSTNGIKLILFSKKEIC 735
 DB 683 -mflvstlrvpk---vasahlegylneysltsmknhsllslg 723

RESULT 10

AAE07901 standard; Protein; 1032 AA.

AAE07901;

01-NOV-2001 (first entry)

C. botulinum C2 translocation domain with Bont/F-binding domain #2.

Neuronal cell; binding domain; translocation domain; stroke; epilepsy;

tumour; infection; neurodegenerative disease; gene therapy;

botulinum neurotoxin type F; Bont/F.

Clostridium botulinum.

NO200158936-A2.

16-AUG-2001.

04-DEC-2000; 2000MO-GB04644.

02-DEC-1999; 99GB-0028530.

07-APR-2000; 2000GB-0008658.

(MICR-) MICROBIOLOGICAL RES AUTHORITY.

Shone CC, Sutton JM, Silman N;

MP1: 2001-514643/56.

Example 2; Page 48; 50pp; English.

The invention relates to a non toxic polypeptide, for delivery of a
 CC therapeutic agent to a neuronal cell, which comprises a binding domain
 CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
 CC as HC) that binds to the neuronal cell and a translocation domain (amino
 CC terminal half of HC, designated as HN), that translocates the therapeutic
 CC agent into the neuronal cell, where the translocation domain is not a HN
 CC domain of a clostridial neurotoxin and is not a fragment or derivative of
 CC a HN domain of a clostridial toxin. Polypeptides of the invention are
 CC useful for the treatment of a disease state associated with neuronal
 CC cells. The polypeptide constructs are useful for delivering therapeutic
 CC substances to neuronal cells. They are useful to treat disorders of the
 CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
 CC and infection. They are also useful in gene therapy. The present sequence
 CC is C. botulinum C2 enterotoxin translocation domain with botulinum
 CC neurotoxin type F (Bont/F) binding domain used in the exemplification of
 CC the invention.
 XX
 XX Sequence 1032 AA;

Query Match 22.6%; Score 854; DB 22; Length 1032;
 Best Local Similarity 32.8%; Pred. No. 1.2e-48;
 Matches 247; Conservative 139; Mismatches 259; Indels 108; Gaps 31;

QY 3 KOENRLNSES--SSQGLGYFSDLPQAPWYSSSTGDSIPSELENIPEENQYF 60
 DB 4 klenysvknknyftinglmgyfien-dfnlnlspcltgnlftskedinslg-nkll 61
 QY 61 QSATMSGPIKVKKSDERTFATSADN-HYTMWVDQEVYKASN-SNKIRLEKGLYQIKI 118
 DB 62 ksarwagllkpslftgyelstnspcrvel--ngelfnlstntsvnlhlgnydydlrl 118

```

OY 119 -OYORENPTKGLD-FKLYTDSOKKEVYSSDNLQELKOKSSNRKRKSTAGPTV- 175
DB 119 eqlnsenqllknyegiklywetadllikeilpsevl---lkpysntnekskifpnltf 174
OY 176 -----PDRNDGIPDSLEVEGYVDVKNKRFELSPWISNIHEKKGLTKRKSSPEK 225
DB 175 snaklkanaanrtdtdgldpewelnqlytmqkavawdkfaan-----gykkyvsnpdk 229
OY 226 WSTASDPYSDPEKVTGRIDKNVSPPEARHPVAAPYIVHDMENILISKNEQOSTONTDSE 285
DB 230 pclandpdyldfekvsgldpsvsmvardpmisaypivgmerlvyskse-----tlfgds 285
OY 286 TRTSKNTSTSRTHSEVHGNAEYHASFEDIG-----SVSAGFSNSNSTVAIDHSL 338
DB 286 tkmskstsstlnltv---gaevsgldagglfprfmsasanyshltwqstvdct- 342
OY 339 SLAERTWAEVGLMTADTARLNANRYVNGTAPYIVNLPPTSLVLRKNOTATIKAKE 398
DB 343 --tge-sfsgslntgesaylmpnlrlyntgclapynvcpdltyldk-qvavtlkgge 398
OY 399 NQLSQILAPNNYPSKNIAPALNADODSSPTTMYNOFLEKTKQLRLDTQVYGN 458
DB 399 slldqylnpggtyplgeppmahtmdqtsrllpdyngqlksldngvtmstsgltgn 458
OY 459 IATYFENGVRVDTGSMSEVLPQIOETTARIT-FNGKDLNVERRIAANPSDPLET 516
DB 459 fakyn-sngnlvld-gnmwpgylytlkstaletstsgltqva---vvaaptsdpdk 513
OY 517 TKPDMTLEKALIAFGNEPNMGLOYOGDIT---EPDFNFDOOSTONIKNOELMANTN 573
DB 514 t-pltcltegalkafalekngkfyfhgletskneklyqvlndentfenglntadkd 572
OY 574 IYTVLDIKILKNKNIILIRDRFRHYDRNNIAGVADSEVYEAHREYRSTEGLLNIDK 633
DB 574 lmbel--lkhn--tmllkvyltf---kenls-----sintmgyndkllilyfuk 615
OY 634 DIRKILSGYVEIEDETEGLKEVINDRYDMANTSSLRDCKTFIDFKKY-----NDKLEP 687
DB 616 lykkikhsnld-----myeanckfidlagysnlsangvyl 653
OY 688 YISNPNKVVAVYAKENTITINSENGDRTSNG 720
DB 654 ysturn-qfgly--askpsevnlaqndllyng 683

RESULT 11
AAE07900
ID AAE07900 standard; Protein: 1092 AA.
AC AAE07900;
DT 01-NOV-2001 (first entry)
DE C. botulinum C2 translocation domain with BONT/F-binding domain #1.
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy;
KW botulinum neurotoxin type F; BONT/F.
OS Clostridium botulinum.
PN MO200158936-A2.
PD 16-AUG-2001.
PE 04-DEC-2000; 2000MO-G804644.
PR 02-DEC-1999; 99GB-0028530.
PS 07-APR-2000; 2000GB-0008658.
PA (MICR-) MICROBIOLOGICAL RES. AUTHORITY.
XX

```

```

PI Shone CC, Sutton JM, Silman N;
XX WPI: 2001-514643/56.
XX
DR New non toxic polypeptide for delivery of a therapeutic agent for the
XX treatment of a CNS disorder comprising a binding domain that
XX translocates the therapeutic agent into the neuronal cells -
XX Example 2; Page 47; 50pp; English.
XX
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC as HC) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides associated with neuronal
CC are useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for treating disorders of the
CC CNS including neurodegenerative diseases; stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is C. botulinum C2 enterotoxin translocation domain with botulinum
CC neurotoxin type F (BONT/F) binding domain used in the exemplification of
CC the invention.
XX
SQ Sequence 1092 AA;
XX
Query Match 22.5%; Score 850; DB 22; Length 1092;
Best Local Similarity 32.0%; Pred. No. 2; 3e-48;
Matches 249; Conservative 150; Mismatches 280; Indels 100; Gaps 33;
OY 3 KOEURLNSES--SSGILGYFSDLNFOAPVNTSSTGDLSPSELENIPSNQYF 60
DB 4 kfensvsnknkyffinglmglyfen-dfnlnlspldgnltksednslly-nkrl 61
OY 61 OSAMSGEIKVKKSDSEYTFATASN-HVTMWVDDQEVYKASN-SNRTRIEKRLYQRI 118
DB 62 ksarvlgllkpsltgelylstenpncrvel---ngelfnlstnhtcnvnlignvydrl 118
OY 119 -OYORENPTKGLD-FKLYTDSOKKEVYSSDNLQELKOKSSNRKRKSTAGPTV- 175
DB 119 eqlnsenqllknyegiklywetadllikeilpsevl---lkpysntnekskifpnltf 174
OY 176 -----PDRNDGIPDSLEVEGYVDVKNKRFELSPWISNIHEKKGLTKRKSSPEK 225
DB 175 snaklkanaanrtdtdgldpewelnqlytmqkavawdkfaan-----gykkyvsnpdk 229
OY 226 WSTASDPYSDPEKVTGRIDKNVSPPEARHPVAAPYIVHDMENILISKNEQOSTONTDSE 285
DB 230 pclandpdyldfekvsgldpsvsmvardpmisaypivgmerlvyskse-----tlfgds 285
OY 286 TRTSKNTSTSRTHSEVHGNAEYHASFEDIG-----SVSAGFSNSNSTVAIDHSL 338
DB 286 tkmskstsstlnltv---gaevsgldagglfprfmsasanyshltwqstvdct- 342
OY 339 SLAERTWAEVGLMTADTARLNANRYVNGTAPYIVNLPPTSLVLRKNOTATIKAKE 398
DB 343 --tge-sfsgslntgesaylmpnlrlyntgclapynvcpdltyldk-qvavtlkgge 398
OY 399 NQLSQILAPNNYPSKNIAPALNADODSSPTTMYNOFLEKTKQLRLDTQVYGN 458
DB 399 slldqylnpggtyplgeppmahtmdqtsrllpdyngqlksldngvtmstsgltgn 458
OY 459 IATYFENGVRVDTGSMSEVLPQIOETTARIT-FNGKDLNVERRIAANPSDPLET 516
DB 459 fakyn-sngnlvld-gnmwpgylytlkstaletstsgltqva---vvaaptsdpdk 513
OY 517 TKPDMTLEKALIAFGNEPNMGLOYOGDIT---EPDFNFDOOSTONIKNOELMANTN 573
DB 514 t-pltcltegalkafalekngkfyfhgletskneklyqvlndentfenglntadkd 572

```

OY 574 IYVLDKIKLNKNNILIRDKRFRHYDRNNIAGVD-----ESVKEAHR-----VINS 623
 Db 573 Imhcl-1krn--mlllvkyltkenisslnldnrfvgymclntrskvgdglvraa 628
 OY 624 TEGLLNIDKDIR-----KILSGYIEIEDTGLKEVINDRYDL-----NISS 667
 Db 629 ttafsfk-skelkypgyymtrfivqsyepflmsyndklllyfnklykklkdnslld 687
 OY 668 IRODKRTFIDPKRY-----NDKLPYISNPNRYKVNVAVENTIINPSENGTSTNG 720
 Db 688 mysnkfkldisgysnslngvlysturn-qfgyly--sakpsvnaqdndillyng 743
 RESULT 12
 ID AAE07902 standard; Protein; 1112 AA.
 XX AAE07902;
 AC AAE07902;
 XX 01-NOV-2001 (first entry)
 DE C. botulinum C2 translocation domain with TeNT binding domain #1.
 XX
 KM Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
 KM tumor; infection; neurodegenerative disease; gene therapy;
 KM botulinum neurotoxin; tetanus neurotoxin; TeNT.
 OS Clostridium botulinum.
 OS Clostridium tetani.
 XX WO200158936-A2.
 XX 16-AUG-2001.
 PD 04-DEC-2000; 2000MO-GB04644.
 PF 02-DEC-1999; 99GB-0028530.
 PR 07-APR-2000; 2000GB-0008658.
 XX
 PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
 PI Shone CC, Sutton JM, Silman N;
 XX WPI: 2001-514643/56.
 DR
 XX New non toxic polypeptide for delivery of a therapeutic agent for the
 PT treatment of a CNS disorder comprising a binding domain that
 PT translocates the therapeutic agent into the neuronal cells -
 PS Example 2; Page 49; 50pp; English.
 CC The invention relates to a non toxic polypeptide, for delivery of a
 CC therapeutic agent to a neuronal cell, which comprises a binding domain
 CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
 CC as Hc) that binds to the neuronal cell and a translocation domain (amino
 CC terminal half of HC, designated as HN), that translocates the therapeutic
 CC agent into the neuronal cell, where the translocation domain is not a HN
 CC domain of a clostridial neurotoxin and is not a fragment or derivative of
 CC a HN domain of a clostridial toxin. Polypeptides of the invention are
 CC useful for the treatment of a disease state associated with neuronal
 CC cells. The polypeptide constructs are useful for delivering therapeutic
 CC substances to neuronal cells. They are useful to treat disorders of the
 CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
 CC and infection. They are also useful in gene therapy. The present sequence
 CC is C. botulinum C2 enterotoxin translocation domain with tetanus
 CC neurotoxin (TeNT) binding domain used in the exemplification of the
 CC invention.
 XX
 SQ Sequence 1112 AA;
 Query Match 22.3%; Score 841.5; DB 22; Length 1112;

Best Local Similarity 32.1%; Pred. No. 9e-48;
 Matches 240; Conservative 143; Mismatches 269; Indels 95; Gaps 30;
 OY 3 KOENRLNSES--SSOGLGYFSDNLFOAPMYVSTTGDS--ESSELENIPSENOYF 60
 Db 4 kfevsynskkyftfinglyfien-dfflnlspfldgnllfskedsnllg-nkll 61
 OY 61 QSAIWSGFIVKKSDEYFATSADN-HYIMVWDQEVINKASN-SNKIRLEKRLQIKI 118
 Db 62 ksarwlgllkptltgeylstnpsncrvel--pgelfnlstlnstnyllqgnvydirl 118
 OY 119 -QYREMPTEKGD-FKLYTDSOKNKEVISDMLOPELKOK--SNKRSKRSAGPYR 175
 Db 119 eqlmsengllkoyegklywetsdlkelpsevl--lkpn/sntneksktpnllf 174
 OY 176 -----PRDNDGIPDSLEVEGYVDVKNKRTFLSPMISNHEKGLTKYKSPER 225
 Db 175 snalkkananrdtdgldpeweinglyymqkavaddkfaa-----gykkyvsnplk 229
 OY 226 WSTASDPYSDFEKYGRIDKNVSPENRHPVLAATPIYHDMENITLSKNEOSTOTDSE 285
 Db 230 pclandpytdfekvsqgldpsvswardpmlsapplyvgwmerlvwskse-----ltgds 285
 OY 286 TRTISKNTSTSRHTSEVHGNAEVAHASFEDIG-----SVSAGFSNNSSTVAIDSL 338
 Db 286 tkmskstkshstlnlv--gaevsqslqslgdlfpfmsaslnyshwtqntstvdtd- 342
 OY 339 SIAGERTWAEIAGLNTADTARLNANIRYVNTGAPINVLPTTSVZGKNQTLATIKAE 398
 Db 343 -tge-sfsgslstlgesaylmpnlryngtapyvnpitllvldk-qsvatlqge 398
 OY 399 NQLSQIAPNNYVPSKNIAPIALNADDFSSPTIMNYNOFLE-EKTKQLRLTDDVOYN 458
 Db 399 slldgylnpggytprlgeppmalntndgsrllipnqlksidngtvlatsqftgn 458
 OY 459 IATNFNGRVRVDGSMWSEVLPJOETARTII--ENGKDLNVERIAVAPSPLEET 516
 Db 459 fakyn-englvtld-gnnwgpjlygtlktstasllfsfggtqva--vvaqpfsgpek 513
 OY 517 TRPMTKEKALIAFGFENPENGMLQYOGKDIR--EFDNFPOQ/SONIKNOIAELMNTN 573
 Db 514 t-pkltlegalkatalekknkfyfnglsksnekqyflidsrtnndfengklnadd 572
 OY 574 IYVLDKIKLNKNNILIRDKRFRHYDRNNIAGVDESIVKEAHRVINSSTBGLLNDK 633
 Db 573 Imhcl-1krn--mlllvkyltkenisslnldnrfvgymclntrskvgdglvraa 608
 OY 634 DIRKILSGYIEIEDTGLKEVINDRYDLNINISSLRDGR--FTI--DRKYNKLPY 688
 Db 609 ---qsmtglstrskvgdglvraaatafsfkskelkypgyymtrfivqsyepflkndcw 665
 OY 689 ISNPNRYKVNVAVENTIINPSENGD 715
 Db 666 vdn---eedldvllkksllnldind 689
 RESULT 13
 ID AAM60224 standard; Protein; 880 AA.
 XX AAM60224;
 AC AAM60224;
 XX 28-SEP-1998 (first entry)
 DE Bacillus thuringiensis insecticidal toxin 177C8.
 DE Insecticide; pesticide; toxin; delta-endotoxin;
 KW biological control; lepidopteran; coleopteran.
 XX
 OS Bacillus thuringiensis strain PS17C8 (NRRL B-21667).
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 253

FT		/note= "encoded by YTA"
FT	Misc-difference	675
FT	Misc-difference	/note- "encoded by AC"
FT	Misc-difference	846
FT		/note- "encoded by RAA"
PX		
PN	W09818932-f2.	
PD	07-MAY-1998.	
PF	30-OCT-1997;	97WO-US19804.
PR	30-OCT-1996;	96US-0029848.
PA	(MYCO) MYCOGEN CORP.	
PI	Dullum CJ, Feltelson JS, Loewer D, Muller-Cohn J;	
PI	Narya KE, Schmelts JL, Schnepf HE, Schwab G, Stamp L;	
PI	Schnepf BA;	
XX	WP1; 1998-272226/24.	
DR	N-Psdb; AAV30307.	
PT	Bacillus thuringiensis isolates - used for producing pesticidal	
PT	toxins and nucleotide sequences for control of lepidoptera and	
PT	coleopterans	
XX		
PS	Claim 5; Page 81-84; 13pp; English.	
XX	This polypeptide comprises a novel soluble toxin of Bacillus	
CC	thuringiensis (B.t.) strain PS177C8 (NRRL B-21867). The toxin	
CC	belongs to a novel family of B.t. toxins that have toxicity	
CC	against non-mammalian pests. Its amino acid sequence was deduced	
CC	from a novel DNA fragment (see AAV30307) obtained by PCR from	
CC	cellular genomic DNA of PS177C8 disclosed and claimed are novel	
CC	B.t. isolates and toxins (see AAW60218-32) that have activity against	
CC	lepidopteran and/or coleopterian pests, isolated genes, probes	
CC	and primers (see AAV30288-321 and AAV9734-87) useful for production	
CC	of the toxins and for the identification and characterisation of	
CC	these toxins, and transformed hosts, particularly plant and	
CC	bacterial hosts. The invention provides 8 entirely new families of	
CC	toxins from B.t. isolates. The toxins have the additional ability	
CC	to form pores in cell membranes, and can be used to facilitate	
CC	entry of a second agent into a target cell.	
XX		
XX	Sequence 880 AA:	
XX		
XX	Query Match	21.0%; Score 792.5; DB 19; Length 880;
XX	Best Local Similarity	30.5%; Pred.No.1.2e-44;
XX	Matches 246; Conservative 134; Mismatches 298; Indels 129; Gaps	34;
OY	4 OENLNHESSSSOGGLAGFFSOLFNOAPMWTVSSTGDLSPSSLEN--IPSENQFO 61	
OY	: : : : : : : : :	
OY	39 qkng--qkemarklglyfyfkgtf-snlmefaprtelliydqtanlkldkkqgyg 94	
OY	62 SAIVSGFIRKKSDEYPFAPSANHNHVMWDVDFINRKASNSNKRILEKGRLYOIKRIOYO 121	
OY	: : : : : : :	
OY	95 slrilyglsketgdtfnlsedeqallnqtklankkkykvahlekylpykleiyg 154	
OY	122 RENTKEGID-----PLTWDSOMKKEVISNDLDELKOS-----N 162	
OY	: : : : : : :	
Db	155 sd--tknidstktelkrlkfidsqngpqyqgdelpmfefnrkkesgeqlafkpsklntf 212	
OY	163 SRRKRSISAGFPVPVRONDGIPDLSLEVCGTYVDVKKKRFLSPWISNIHKKGGLTKRYSS 222	
Db	: : : : : : :	
OY	213 qkmrteide----ctcdgdslplwengyytl---gnrlavkwddsl-asqykylvsn 264	
OY	223 PEKMSTADSYSDFEKYVGRIDKNVSPAARHPLAAYPIVHVDMENILISKMDOSTONT 282	
Db	: : : : : : :	
OY	265 pleshtvgdpdydvycakardidalsnaketlnptlvaaftsvpnsmekyilspeants--- 320	
OY	283 DSETRITSNKTSTSRHTSEVHONAEVASPFDIGSVSAGFSUNSSN---TVAIDHSLS 359	

Db 321 -----nsveshstsnrcyft-----nte-----gastveagigckgkistfgsvnyqhsbet 363

Qy 340 LGEETVAEEMG-----LNTADTARLNNINRYVETGTAIRYNYLCTGSLVIGKNOTLAFIK 395

Db 364 vqge--vqstugntsgfntaasagynnanvinnvggaayvkrkcsfvi--ndtlatlc 420

Qy 396 AKENQSLQIAPNNNYPPSKNNAPIALNAQDQSSSTPTMTNNTQGLEKTKQLRLDDQV 455

Db 421 aasnsalanpspegyppkxgqnglatfsmddtnshphtlcnkkyvdlinnkmmletnqt 480

Qy 456 YQNTATYFENGVRVDGSGMSWEVLQIQETARTIENGKOLNVERRIAAVNSDPLE 515

Db 481 dg---vylktdtgnvntvgewngvliqikakastlvdgge--rvaekvaakdyemped 536

Qy 516 TRKPDMLKEALKIAR--GFENPNQNTLOYGKDITEFDF--NFQDQTSNIRKNQDAEL-- 569

Db 537 kt-paltlkdaalhsypdeakeleglllyknkpiyassvmyclidentakavckgindctg 595

Qy 570 ---NANTYIVYLKIKLTKNNKILRLDRKRFHYDRNNIAVGADESYVKEAAREVINSSTEG 626

Db 596 kEKdVshlydv---kltpckmvtlk--lsillydn--aesdnshgkvntnlivsgang 647

Qy 627 -----LLMLID-----KDIRKILSGYVIEIDPE-----GLKEY 655

Db 648 kkygsnnpdantllntdeqeklnkrystlmyksekntgcetallidgelypiltctkvv 707

Qy 656 INDRDMLNIT--SLRQDQKTPIDFKKYNKLPDYLSNPNNKAVAVATKENTILNPSN 713

Db 708 ndnykridliahnkspisshikt-ndetlflfddisi--lwaakpen--ltdeei 763

Qy 714 GDT-STNGIK---KILIPSKKEYEIGZ 736

Db 764 kqlsyryikldeglilldkkgyihye 790

RESULT 14

AAI95277

ID AAI95277 standard; Protein: 861 AA.

XX AC AAI95277;

Db DT 18-APR-2000 (first entry)

XX DE MIS toxin from B. thuringiensis strain PS177C8.

KW Bacillus thuringiensis; toxin; endotoxin; pesticiida; plant pest;

KM lepidopterans; cleopterans.

XX OS Bacillus thuringiensis.

XX FH Key Location/Qualifiers

FT Misc-difference 846

FT /note= "unknown"

XX PN MO9957282-A2.

XX PD 11-NOV-1999.

XX PE 06-MAY-1999; 99MO-US099997.

PR 06-MAY-1998; 98US-0073898.

XX RA (MYCO) MYCOGEN CORP.

XX PEitelson JS, Schepf HE, Narva KE, Stockhoff RA, Schmelts J;

PI Loewer D, Dullum CJ, Muller-Cohn J, Stamp L, Wottrill G;

PI Flinstad-Lee S;

XX DR WPI: 2000-096811/08.

XX N-PSDB: AA258786.

XX New polynucleotides encoding pesticidally active proteins, useful for

PT transforming plants for controlling pests
XX
PS Disclosure; Page 68-71; 104pp; English
XX

CC The invention relates to novel *B. thuringiensis* isolates, and genes
CC encoding pesticidal toxins which are toxic to non-mammalian pests. The
CC genes are useful in the control of non-mammalian pests and especially
CC plant pests (e.g. lepidopterans and/or coleopterans). The polynucleotides
CC are useful for transforming plants for controlling plant pests; for
CC designing plasmids and probes useful for the identification and
CC characterization of genes that encode pesticidal toxins. The present
CC sequence represents a *B.t.* toxin.

50 Sequence 881 AA;

Query Match	21.0%	Score 792;	DB 21;	Length 881;
Best Local Similarity	30.4%;	Pred. No. 1.4e+44;		
Matches 246;	Conservative 135;	Mismatches 297;	Indels 130;	Gaps 34;

[illegible]

```

Db      764      1 kqysrygkledgilldkkgihyge 791
          |  |||  |||  |  |  :

```

RESULT 15
APR01220

ID AAR91239 standard; Protein; 884 AA.

AC AAR91239;

DT 14-AUG-1996 (first entry)

DE B. cereus VIP1A(a) insect-specific protein

KW Pesticide; insecticide; biological control agent; Lepidoptera;

KW western corn rootworm; *Diabrotica virgifera virgifera*; VIP.

.05 *Bacillus cereus* strain AB78 (NRRL B-21058)

Key	Location/Qualifiers
FH	

FT	1:03	Sig_peptide
FT	1:03	Sig_peptide

AA W09610083-A1.
PN

PD 04-APR-1996.

AA 27-SEP-1995; 95WO-EP03826
PF

XX 05-JUN-1995: 95US-0463483
PR

PK 28-SEP-1994; 9405-0314594
XX

PA (CIBA) CIBA GEIGY AG
XX

PI carr B, Desai NM, Duck NB, Estruch JJ, Kostichka K,
PI Koziel MG, Mullinsma, Nye GT, Warren GW:

XX
DR WPT: 1996-200921/20.

DR N-PSDB; AAT13940.
XX

PT Bacillus strain producing insecticidal protein during vegetative growth - used in the control of lepidoptera and Coleoptera pests

XX :
PS Claim 15: Page 121-124: 242no: English

XX Insect-specific protein YTP1A(a) (AAP91239) of *Bacillus* AB78

CC shows activity against *Diabrotica* spp. pests such as the
CC western corn rootworm. It is encoded by the *WDR14* gene (AAT13040).

CC isolated from a cosmid clone of AB78. VIP1A(a) can be expressed in CC or bacterial hosts to provide biological control agents having

CC increased activity or target range, or can be expressed in transgenic plants can make to improve forest productivity.

CC expressed as a fusion protein (see also AAR91245) with auxiliary protein
CC (AAR91245) (AAR91245)

XX
XX

Query Match

Matches 248; Conser

Query Match	20.98;	Score 790;	DB 17;	Length 884;
Best Local Similarity	30.48;	Pred. No. 1.9e-44;		
Matches 248;	Conservative 130;	Mismatches 292;	Indels 146;	Gaps 34

```

0Y      4 0ENRLNLSSESSOGSLGTYFPEDLNFQAPMWTYSSTTGLSPSELLEN--ISENQDYO 61
      1 1 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      42 qnq--qkmdrktglllyfryfgkdf-snlmtfaprrsttlllydqtentkllkqgey 97

0Y      62 SAIVSGCFKIVKSPETPTPTASDNIHVMWMDQEVINKNSNKNKRLREKGLQVOTIYO 122
      1 1 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      98 stwvlgllqsketgdfvfnlsedegallnqgklsnkbkqevnhlekglvpikley 15

0Y      122 RNPPEKGLD-----FKLYMTDSQNKKEVSSDNLQLEELKOKS-----N 167
      : 1 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 158 sd--tkfnidsktfkellfkidsqnpqvgdelirnefnkkesgeflakpskinlft 215
QY 163 SRKRSTSAQPIPPDRNDGIPDSLEBYVDYVKNKRFPLSPMISNIHEKKGLTYKSS 222
Db 216 qkmkreided--tdtdgdspdlweenytl----gnriavwddsl-askytlkfvan 267
QY 223 PEKSTASDPYSPDKVGRIDKNVSPARHPLAAPIVHVMENITLSKNEOSTONT 282
Db 266 pleshlygdycdyekaardidlnakefnplvaafpsvnmekvnlspnmls---- 323
QY 283 DSERTISKMTSRTKRTSEVHGNAAVHASFPIGGVSAGFSNSNS--TVAIDHSLS 339
Db 324 -----nveshstnswyt-----nte-----gasveaglgpkylsfyvsvnyghset 366
QY 340 IAGERTVAETMG-----LNTADPRLNANTRYVNTGAPRYNVLPPTSVLGKNOTLATIK 395
Db 367 vage--wgtstqntsqfntasagylaanrynvgtgalydvkptstfvl-ndtlatit 423
QY 396 AKENQLSQIIAPNNYPSKNIAPIALNAODFSSPTMANYNOFLELEKTKOLRLPTDOV 455
Db 424 aksnstalnispgeaypkkgngjaltsmddfushpiltlnkkqvdllnnkpmmletngt 483
QY 456 YGNATYFNFGNVRVDRGNSWSEVLPOIOETTARITFNGKDLNVERIAVANPSPDLE 515
Db 484 dg---vyklldhgnlvrgewngvlgjakaktasllvdage-ryaekrrvaakdyenped 539
QY 516 TTKPDWTKFAKIAF--GFNEPENGLOYOGKDTFEDF--NFDQTSQNIKNQLAEL-- 569
Db 540 kt-pelikaalkleydelkelegillyknkplyessvmcyldentakvckqjndltg 598
QY 570 ---NATNITYVLDTKLAKAKNLTILNDKRFHYDRNNIAGADESVVKAHREVIANSSTEG 626
Db 599 kfkdvshlydv---klcpkmvrlk-lsillydn--aesndnsigkwtlnlvsgngng 650
QY 627 -----LLIND-----KDRIKILSGYVEIEDTE-----GLKE 654
Db 651 kkyssnnpdanltnldageklnkrdyislymkseknitgceitldageilyptlctvn 710
QY 655 VINDRYMLN-----ISSLRDQKTFIDPKKYNDKLPYISNPNYKVNVYAVTKE 704
Db 711 vnkdnykrldlahnknapslsh-----ikndeitlfdwdisl-tavasikpe 760
QY 705 NTIINPSENGDT-STNGIK--KILIFSKGYEIGZ 736
Db 761 n--ltdseikqlysrjgikledgilldkkgjlyhyge 794

Search completed: October 4, 2002, 14:35:58
Job time: 103 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: October 4, 2002, 14:34:15 ; Search time 24.37 Seconds

(without alignments)
2902.000 Million cell updates/sec

Title: US-09-848-909-21

Sequence: 1 EVKQENRLNLSSESSSQGL.....STNGIKILIFSKKGYEIG 736

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3774	99.9	764	2 139934	protective antigen
2	864	23.4	875	2 140862	lotia toxin compo
3	235.5	6.2	192	2 139933	cryptic protein -
4	235.5	6.2	204	2 G59104	hypothetical prote
5	199	5.3	4688	2 F82885	hypothetical prote
6	186	4.9	1227	2 C97033	uncharacterized pr
7	184.5	4.9	2401	2 T28676	rhoxy protein -
8	183.5	4.8	2529	2 B64635	toxin-like outer m
9	183	4.8	4152	2 T31102	filamentous hemag
10	182.5	4.8	1125	2 E90598	membrane nucleas
11	181	4.8	2269	2 T28677	rhoxy protein -
12	179.5	4.8	1639	2 S05603	major merotoxite su
13	178.5	4.7	1072	2 A86827	hypothetical prote
14	178	4.7	752	2 G90589	hypothetical prote
15	178	4.7	1302	1 J06009	surface-located me
16	178	4.7	1658	2 S55101	hypothetical prote
17	178	4.7	2178	2 S55805	alpha-toxin - Clos
18	178	4.7	4913	2 T11105	hypothetical prote
19	178	4.7	6713	2 B89921	hypothetical prote
20	176	4.7	1193	2 S68218	hemolysin (impor
21	176	4.7	1635	2 A10452	hypothetical prote
22	173.5	4.6	821	2 S67087	hypothetical prote
23	173.5	4.6	1802	2 S52611	repeat organellar
24	172	4.6	1939	2 T18372	hypothetical prote
25	171.5	4.5	1365	2 T30822	hypothetical prote
26	171.5	4.5	2399	2 H71879	toxin-like outer m
27	171	4.5	1465	2 S31262	toxin protein - year
28	171	4.5	1803	2 S56894	toxin protein - year
29	170	4.5	1837	2 T41023	probable nuclear p

30	169.5	4.5	786	2 T18469	hypothetical prote
31	169.5	4.5	1230	2 S56850	SMC1 protein homol
32	169.5	4.5	2340	2 B71704	cell surface antig
33	169	4.5	1308	2 E71622	probable membrane
34	168	4.4	1631	1 SAZOK1	major merotoxite su
35	168	4.4	5005	2 F82884	hypothetical prote
36	167.5	4.4	1051	2 T18351	hypothetical prote
37	167	4.4	1315	2 T28679	fibriogen-binding
38	166	4.4	3724	2 T18427	hypothetical prote
39	165.5	4.4	979	2 J00894	hypothetical prote
40	165.5	4.4	1115	2 T41342	probable coiled-co
41	165	4.4	3216	2 C90538	hypothetical prote
42	164.5	4.4	1467	2 PC1253	hypothetical prote
43	164	4.3	1467	2 F82870	tye protein - year
44	164	4.3	1516	2 E71619	serine proteinase
45	163.5	4.3	624	2 PC6003	surface membrane p

ALIGNMENTS

```

RESULT 1
139934
C:Species: Bacillus anthracis
C:Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 01-Dec-2000
C:Accession: I39934; S69160; F59104
R:Melkos, S.L.; Lowe, J.R.; Eden-McCutchan, F.; Vodka, M.; Leppa, S.H.; Schmidt
Gene 69, 287-300, 1988
A:Title: Sequence and analysis of the DNA encoding protective antigen of Bacillus
A:Reference number: I39933; MUID:8912073
A:Accession: I39934
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-764 <RES>
A:Cross-references: GB:M22589; NID:q143280; PIDN:AAA22637.1; PID:q143282
R:Frederman, I.C.; Gordon, V.M.; Leppa, S.H.; Klimpel, K.R.; Birch, N.P.; Loh, Y.
Arch. Biochem. Biophys. 316, 5-13, 1995
A:Title: In vitro processing of anthrax toxin protective antigen by recombinant p
A:Reference number: S69160; MUID:95142670
A:Accession: S69160
A:Molecule type: protein
A:Residues: 197-202 <FRI>
R:Okinkwa, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Kelm, P.;
J. Bacteriol. 181, 6509-6515, 1999
A:Title: Sequence and organization of pXOI, the large Bacillus anthracis plasmid
A:Reference number: A59091; MUID:9945483
A:Accession: F59104
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-313, 'O', 315-764 <OKI>
A:Cross-references: GB:AF065404; NID:q4894216; PIDN:AO32414.1; PID:q4894326
A:Experimental source: strain Sterne
A:Note: similar to anthrax toxin moiety, protective antigen, pagA formerly pag, p
C:Genetics:
A:Gene: pXOI-110
A:Genome: plasmid
C:Function:
A:Description: three component exotoxin; protective antigen binds to receptors on
y active components edema factor or lethal factor; the complex is internalized by
C:Keywords: exotoxin
F:1-29/Domain: signal sequence #status predicted <Sig>
F:30-196/Domain: propeptide #status predicted <Pro>
F:197-202/Product: protective antigen #status experimental <Mat>

Query Match 99.9% Score 3774; DB 2; Length 764;
Best local similarity 100.0% Pred. NO. 1.6e-183; Indels 0; Gaps 0;
Matches 735; Conservative 0; Mismatches 0;
DB 30 EVKQENRLNLSSESSSQGLGYTFSDFNQAPVYTSSTTGDLSPPSELENIPSENOYF 60
1 EYKQENRLNLSSESSSQGLGYTFSDFNQAPVYTSSTTGDLSPPSELENIPSENOYF 60

```

[illegible]

Query Match 6.2%; Score 235.5; DB 2; Length 192;

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: C97033
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1127 <KOR>
 A:Cross-references: GB:AE001437; PIDN:NAK79054.1; PID:915023995; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetic: CAC1080

Query Match 4.9%; Score 186; DB 2; Length 1227;
 Best Local Similarity 19.0%; Pred. No. 0.12;
 Matches 167; Conservative 16; Mismatches 287; Indels 288; Gaps 44;

```

37 SSTGDISPSELEINPSENOYFOSAIY-SGFTKKKSDYETFTASDNHYTMVDDOE 95
14 SVTTSVALLSKRAIKAVAAADSVKSSISNSNINVGKE-----VOEKE 58
96 V-INKASNSKRIKLEKRLVQIKIYORENPTER-----GLDFKLYTMDSONKKEVI 146
59 YVWKNKNDNS- KVSSENQVS---NKNSNPKVSSSEIOSINKVNLQVQNNKSVL 114
147 SSDNLOIPELKOKSSNSRRK-----RSTSGPTVPDNDGIPDSLEYETVDYNNKT 201
115 AASNVDEVKAKKINSQVOTSDYIAIGETVYKEDLDLIINNAIVDASAG--TDLSEVEI 172
202 F-----LSPWISNI-----HEKGL-LTKYK 220
173 YIYISQTAELIQAFRNINDGVANVSDYLLGATFVNDANLDSVKNKYFHKRYATVYK 232
221 SPSKRSKTA-----SDPSPDEKYGRIDKNVSPARHPLVA 258
233 DVATTSALKNNINNGGEGDYTALEVGSGVQPYLD-----VKNKIVYEKO----- 280
259 YPIVAVNMENILISKMEOS--TONTDESPETISK--NTSRTHTSEVHGNVHASFFDI 316
281 -----NKGRLDITELSDASASTIARNTALDN-----MDA 311
317 GGSVSGNSNSSTYVAIDHSL-----SLAGERMAETM-----GLMTADTARLANI 344
312 GAVTLEDYQALGANVPOLAHVADNSLAMDORGVSAIDGINTITITINNIGVGT 371
365 -RYVNTGTAPI-----YVNLPTSL--VLGKQNTLATIKAKENQSLIARNNY----- 411
372 DVIYNSHAVDSNEGNDIDYILNANITIEKTRAKGDL--TIPEVAVNYKEVKTLLDEYNNAA 430
412 -----PSKNLAPIALNADDFSSFTYMNNOFLEKTRKOLRLDTQV----- 455
431 AGOTTIDQYKNDVNAOVODDVAT-----LSMDLKTROCKTILALQDKIDSILNSLN 484
456 ---YGNINATYFENGVRVDTGSMSEVLPOIOETTARI--FENGKDLNVERRIAAVNP 510
485 INSGINIDYD-----KLOTEAVADAKLEAVNDIKIRKNDGRDLTIOETRSVAKT 538
511 SDPLETT-----KPDWTLKEALKIAF--GFENPENGLOYOGKDTTERDENFOOTSNIK-- 563
539 IDYINSTSVKSGDSYDITIGIDGYEI--NIEFVERKEKGGITL--TINIKV 593
564 ---NOLATLN---ATNITYVDL-----KIKLNAMNILLIKRFRHYD 599
594 IEPYQJSEYRIYTGCVGVYDKTLGINNVNOMNIITINELKNKKDYKIDDIOTFVD 653
600 R--NNT---AVGADSEY-----VKAHREYINSSTEGILLMDKDRILKLSGYIE 645
654 NTINNIDVINKIGADAVLSDYFNIGITDYODLLVYVADLKIKNTYKDDVLLIEVEAK 713
646 IEDTEGLK-----EVINDREYMLNISSLRODGKTFIDFKKYNDKPLX-----ISNPK- 695
714 ISSYEALMRINIGEAVIDDFKALGLDI-----NGCLLIATITDLONNKYT 760
696 -----VWY-AVTKENTIIINSENGDTSTNGI 721

```

DB 761 ADEVIARVOAIEIYRALMOIN--LGRATTADYNTLGI 796

RESULT 7

728676
 rhopty protein - Plasmodium yoelli (fragment)

C:Species: Plasmodium yoelli
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #tex_change 01-Dec-2000

C:Accession: Y28676; A45521
 R:Simla, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.

Mol. Biochem. Parasitol. 76, 329-332, 1996
 A:Title: Comparison of two members of a multigene family coding for high-molecular n

A:Reference number: 220507; MUID:97077455

A:Accession: Y28676
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2401 <SIN>
 A:Cross-references: EMBL:U06927; NID:91041784; PID:91041785; PIDN:ABM41263.1

R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.

Mol. Biochem. Parasitol. 42, 241-246, 1990
 A:Title: Identification of the gene for a plasmodium yoelli rhopty protein. Multip.

A:Reference number: A45521; MUID:91101660

A:Accession: A45521

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 2260-2401 <KEE>
 A:Cross-references: GB:M34281

Query Match 4.9%; Score 184.5; DB 2; Length 2401;
 Best Local Similarity 21.6%; Pred. No. 0.39;
 Matches 181; Conservative 120; Mismatches 316; Indels 221; Gaps 43;

```

1 EYKQENRLNSESSESGILGYPEDLN-----FOAPVWYVSTTGDLST 45
402 EYAKENQVQLVKYSNLEIKKHNDQINDIKERAKQVYDOFEKHKMTIPPENMYOK 461
46 PSELEINPSENOYFOSAI--WSGFTKY--KPSDEYFATSADNHYTMVDDOEV---I 97
462 PSELEIKMDE--FLSKYNNKNDPDKYKEVSEHKKFTLNNKITEVSEDEIKYE 518
98 NNASNSKRIKLEKRLVQIKIYORENPTERGLDPLKLYWDSOKKEVSSNLOLPELK 157
519 NFNDSKSLNETK--STEEYQVNTLKKVD--YIKVCLNTNELLITNHNKQTTLK 573
158 OKSSNRK--KRSTSGPTVPDNDGIPD-----SLVEGYVDYNNKRTFLSP 205
574 DKLNQNIKTIKETNSIDKLYTDKRENTLDTKTELETKFTGLSINHESNNKELLTYD 633
206 WISNIHEKGLTYKSSPEKSTASDPSEFKVTRIDKNVSPARHPLVAIYIYVD 265
634 LKANLGRKNEMLYKOFNEK--EKAVEDIKRNDIKRLVSN-----IET 677
266 MENILISKNEODSTONTDESPETISKNTSRTHTSEVHGNVHASFFDIGSAGFS 325
678 IYTSIYINED--TENE--IGKSTELLNTRVLE-----KKNAYT 713
326 NSNSSTVAI--DHSLSLAGERTVAETMGNTADTARLANIRVNTGTAPIYVNLPTSLV 384
714 NMEIKKIKYDQDFGK-----EKNIKYPDEN--KIKNDIDTLNOK 754
385 LGRN--OTLATIKR--KEMOLSOI-----LAAP--NYVP--SKNLAPIA----- 420
755 IDSISTETLEIKSENHIDEIKQIDKLYKVPKTMFNEDEKLEKINIEYERIDKK 814
421 -LNAODFSSPTTMYNOFLEKTRKOLRLDTDOYGNIATYFENGVRVDTGSMNSE 479
815 NIYKEIDKLMNISKIENDKTSLEKKNINLSYKSGISGLIFLOOI--EKKKAKHTIKME 874
480 V---LPOIOETTARITFNGKDLN-----VERRIAAVNSD-----LEFT-----K 518
875 AYIDDIDNKKRSQEI--EKEMNINMDIKMDIKKEMKALNISHDYKIYHTTSKHEK 931

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 14:34:50 ; Search time 13.55 Seconds

(Without alignments)
2103.143 Million cell updates/sec

Title: US-09-848-909-21
Sequence: 1 EVKQENRLNSESSESSQGL.....STNGIKRLIFSKGVEIGZ 736

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Swissprot_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3771	99.8	764	1	PAG_BACAN
2	235.5	6.2	204	1	YPL1_BACAN
3	179.5	4.8	1630	1	MSPI_PLAFC
4	179.5	4.8	1639	1	MSPI_PLAFC
5	178	4.7	1658	1	YMG7_YEAST
6	171	4.5	1803	1	YK13_YEAST
7	169.5	4.3	1230	1	SMC3_YEAST
8	163.5	4.3	2334	1	MAPA_JACSU
9	162.5	4.3	978	1	P115_MCHR
10	162	4.3	1957	1	YD86_SCHRO
11	161	4.3	1790	1	USO1_YEAST
12	160	4.2	1276	1	BKD_CLOBO
13	160	4.2	1385	1	FAP1_SCHRO
14	159	4.2	1000	1	S155_YEAST
15	156.5	4.1	1167	1	CAGA_HELPJ
16	155.5	4.1	1269	1	RBP1_PLAFC
17	155	4.1	1251	1	RBP2_PLAFC
18	155	4.1	1928	1	MYB1_YEAST
19	154.5	4.1	1024	1	RIP3_MOUSE
20	154.5	4.1	1208	1	PCP1_SCHRO
21	154.5	4.1	1726	1	MSPI_PLAFC
22	153.5	4.1	1487	1	YMG7_YEAST
23	153.5	4.1	1487	1	MSD3_YEAST
24	153	4.0	1420	1	SRB3_YEAST
25	152.5	4.0	1116	1	YK54_AQUAE
26	151.5	4.0	1014	1	HEX2_YEAST
27	151.5	4.0	1225	1	Y309_MYCE
28	150.5	4.0	803	1	SMF6_YEAST
29	150.5	4.0	2116	1	MYC2_DICDI
30	149.5	4.0	1271	1	Y338_MYCE
31	149.5	4.0	1356	1	SIR4_YEAST
32	149	3.9	1577	1	HLRA_PROTI
33	148.5	3.9	1233	1	IF16_YEAST

34	148.5	3.9	1460	1	M159_YEAST
35	148.5	3.9	1726	1	MSPI_PLAFC
36	148.5	3.9	1744	1	TANA_XENIA
37	148	3.9	1570	1	P3K1_DICDI
38	147.5	3.9	1091	1	CIC2_RAT
39	147.5	3.9	1093	1	SMI4_YEAST
40	147.5	3.9	2376	1	YIM9_YEAST
41	147	3.9	918	1	YMOB_CAEEL
42	147	3.9	1288	1	VACA_HELPJ
43	146.5	3.9	1162	1	BXEN_CLOBO
44	146.5	3.9	1628	1	NACH_CLOPE
45	146.5	3.9	1916	1	RIP1_YEAST

ALIGNMENTS

RESULT 1	ID	PAG_BACAN	STANDARD:	PRF:	764	AA.
AC	P13423	Q9P5R7	Q9KH69			
DT	01-JAN-1980	(Rel. 13, Created)				
DT	15-OCT-2001	(Rel. 40, Last sequence update)				
DE	01-MAR-2002	(Rel. 41, Last annotation update)				
DE	Protective antigen precursor (PA) (PA-83) (PA83) (Anthrax toxins					
DE	translocating protein) [contains: PA-20 (PA20); PA-63 (PA63)].					
GN	PAG OR PAG OR PXO1-110.					
OS	Bacillus anthracis.					
OC	Plasmid PXO1.					
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;					
OC	Bacillus/Staphylococcus group; Bacillus.					
OX	NCBI_TaxID=1392;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE-09172073; PubMed-3148491;					
RA	Welkos S.L., Lowe J.R., Eden-McCutchan F., Vodka M., Leppla S.H.,					
RA	Schmidt J.J.;					
RT	*Sequence and analysis of the DNA encoding protective antigen of					
RT	Bacillus anthracis.*;					
RL	Gene 69:287-300(1988).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-28_33, BA1024, AND BA1035;					
RX	MEDLINE-99214082; PubMed-10197986;					
RA	Price L.B., Hugh-Jones M., Jackson P.J., Keim P.;					
RT	*Genetic diversity in the protective antigen gene of Bacillus					
RT	anthracis.*;					
RL	J. Bacteriol. 181:2358-2362(1999).					
RN	[3]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-V770-NP1-R / ATCC 14185;					
RX	MEDLINE-20359347; PubMed-1089854;					
RA	Cohen S., Mendelson I., Altboum Z., Kohler D., Jihany E., Bino T.,					
RA	Lettner M., Inbar I., Rosenberg H., Gozes Y., Barak R., Fisher M.,					
RA	Kronman C., Velan B., Shafitman A.;					
RT	*Attenuated nontoxicogenic and nonencapsulated recombinant Bacillus					
RT	anthracis spore vaccines protect against anthrax.*;					
RL	Infect. Immun. 68:4549-4558(2000).					
RN	[4]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-Sterne;					
RX	MEDLINE-99445483; PubMed-10515943;					
RA	Keim P., Koeber T.M., Lamke G., Kumano S., Mahillon J., Manter D.,					
RA	Klein P., Richter D., Stensson R., Jackson P.J.;					
RT	*Sequence and organization of PXO1, the large Bacillus anthracis					
RT	plasmid harboring the Anthrax toxin genes.*;					
RL	J. Bacteriol. 181:6509-6515(1999).					
RN	[5]					
RP	DOMAINS.					
RX	MEDLINE-91332080; PubMed-1651334;					
RA	Singh Y., Kimpel K.R., Quinn C.P., Chaudhary V.K., Leppla S.H.;					
RT	*The carboxyl-terminal end of protective antigen is required for					

P40177 saccharomyc

P04934 plasmodium

O01550 xenopus lae

P54573 dictyostell

P54290 rattus norv

P25302 saccharomyc

P40468 saccharomyc

P34487 caenorhabdi

O92KW5 helicobacte

O06366 clostridium

P26831 clostridium

P29539 saccharomyc

RT receptor binding and anthrax toxin activity.";
 RL J. Biol. Chem. 266:15493-15497(1991).
 [6]
 RN CHARACTERIZATION.
 RC STRAIN-Sterne;
 RA Pubmed-8051159;
 RX Mline J.C., Furlong D., Hanna P.C., Wall J.S., Collier R.J.;
 RT "Anthrax protective antigen forms oligomers during intoxication of
 mammalian cells.";
 RL J. Biol. Chem. 269:20607-20612(1994).
 [7]
 RN CHARACTERIZATION.
 RC STRAIN-Sterne;
 RA Pubmed-11207581;
 RX Beauregard K.E., Collier R.J., Swanson J.A.;
 RT "Proteolytic activation of receptor-bound anthrax protective antigen
 on macrophages promotes its internalization.";
 RL Cell. Microbiol. 2:251-258(2000).
 [8]
 RN TOXIN REGULATION.
 RC STRAIN-Weybridge;
 RA Pubmed-8300513;
 RX Koshier T.M., Dai Z., Kaufman-Yarbray M.;
 RT "Regulation of the Bacillus anthracis protective antigen gene: CO2 and
 a trans-acting element activate transcription from one of two
 promoters.";
 RL J. Bacteriol. 176:586-595(1994).
 [9]
 RN MUTAGENESIS OF PHE-342; PHE-343 AND ASP-344.
 RC STRAIN-Sterne;
 RA Pubmed-7961869;
 RX Singh Y., Klimpel R.R., Arora N., Sharma M., Leppla S.H.;
 RT "The chymotrypsin-sensitive site, PFD315, in anthrax toxin protective
 antigen is required for translocation of lethal factor.";
 RL J. Biol. Chem. 269:29039-29046(1994).
 [10]
 RN MUTAGENESIS OF DOMAIN 4 LOOPS.
 RC STRAIN-Sterne;
 RA Pubmed-10085028;
 RX Varughese M., Texeira A.V., Liu S., Leppla S.H.;
 RT "Identification of a receptor-binding region within domain 4 of the
 protective antigen component of anthrax toxin.";
 RL Infect. Immun. 67:1860-1865(1999).
 [11]
 RN MUTAGENESIS OF TRP-375; MET-379 AND LEU-381.
 RC STRAIN-Sterne;
 RA Pubmed-11178978;
 RX Patra S., Gupta P., Chauhan V., Singh A., Bhatnagar R.;
 RT "Trp 346 and Leu 352 residues in protective antigen are required for
 the expression of anthrax lethal toxin activity.";
 RL Biochem. Biophys. Res. Commun. 281:186-192(2001).
 [12]
 RN MUTAGENESIS OF PHE-581; PHE-583; ILE-591; LEU-595 AND ILE-603.
 RC STRAIN-Sterne;
 RA Pubmed-11554763;
 RX Ahuja N., Kumar P., Bhatnagar R.;
 RT "Hydrophobic residues Phe552, Phe554, Ile562, Leu566, and Ile574 are
 required for oligomerization of anthrax protective antigen.";
 RL Biochem. Biophys. Res. Commun. 287:542-549(2001).
 [13]
 RN MUTAGENESIS OF PRO-289.
 RC STRAIN-Sterne;
 RA Pubmed-11356563;
 RX Khanna H., Chopra A.P., Arora N., Chaudhry A., Singh Y.;
 RT "Role of residues constituting the 2beta1 strand of domain II in the
 biological activity of anthrax protective antigen.";
 RL FEBS Microbiol. Lett. 199:27-31(2001).
 [14]
 RN MUTAGENESIS OF GLN-512; ASP-541; LEU-543 AND ARG-621.
 RC STRAIN-Sterne;
 RA Pubmed-11222612;
 RX Morridge J., Meunier M., Collier R.J.;
 RT "Involvement of domain 3 in oligomerization by the protective antigen
 moiety of anthrax toxin.";
 RL J. Bacteriol. 183:2111-2116(2001).

[15]
 RN MUTAGENESIS OF LYS-426; ASP-454 AND PHE-456.
 RC STRAIN-Sterne;
 RA Pubmed-11113126;
 RX Selman B.R., Nassi S., Collier R.J.;
 RT "Point mutations in anthrax protective antigen that block
 translocation.";
 RL J. Biol. Chem. 276:8371-8376(2001).
 [16]
 RN X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
 RC STRAIN-Sterne;
 RA MEDLINE-97192099; Pubmed-9039918;
 RX Petosa C., Collier R.J., Klimpel R.R., Leppla S.H., Lidington R.C.;
 RT "Crystal structure of the anthrax toxin protective antigen.";
 RL Nature 385:833-838(1997).
 [17]
 RN REVIEW.
 RC STRAIN-Sterne;
 RA Pubmed-11544370;
 RX Mock M., Fouet A.;
 RT "Anthrax.";
 RL Annu. Rev. Microbiol. 55:647-671(2001).
 -1- FUNCTION: One of the three proteins composing the anthrax toxin,
 the agent which infects many mammalian species and that may cause
 death. PA binds to a receptor (Atr) in sensitive eukaryotic
 cells, thereby facilitating the translocation of the enzymatic
 toxin components, edema factor and lethal factor, across the
 target cell membrane. PA associated with LF causes death when
 injected, PA associated with EF produces edema. PA induces
 immunity to infection with anthrax.
 -1- SUBUNIT: Anthrax toxins are composed of three distinct proteins, a
 protective antigen (PA), a lethal factor (LF) and an edema factor
 (EF). None of these is toxic by itself. PA+LF forms the lethal
 toxin (Ltx); PA+EF forms the edema toxin (Edtx). PA-63 forms
 heptamers and this oligomerization is required for LF or EF
 binding. Once activated, at low pH, the heptamer undergoes
 conformational changes and converts from prepore to pore inserted
 in the membrane, forming cation-selective channels.
 -1- SUBCELLULAR LOCATION: Secreted.
 -1- DOMAIN: The molecule is folded into four functional domains. Each
 domain is required for a particular step in the toxicity process.
 Domain 1 contains two calcium ions and the proteolytic activation
 site. Cleavage of the PA monomer releases the subdomain 1a, which
 is the N-terminal fragment of 20-kDa (PA20). The subdomain 1b is
 part of the remaining 63-kDa fragment (PA63). Domain 2 is a beta-
 barrel core containing a large flexible loop that has been
 implicated in membrane insertion and pore formation. There is a
 chymotrypsin cleavage site in this loop that is required for
 toxicity. Domain 3 has a hydrophobic patch thought to be involved
 in protein-protein interactions. Domain 4 appears to be a separate
 domain and shows limited contact with the other three domains. It
 would swing out of the way during membrane insertion. Domain 4
 contains the binding sites for LF and EF and it is required for
 binding to the receptor; the small loop is involved in receptor
 recognition.
 -1- PTM: Proteolytic activation by furin or a furin-like protease
 cleaves the protein in two parts, PA-20 and PA-63; the latter is
 the mature protein. The cleavage occurs at the cell surface and
 probably in the serum of infected animals as well; both native and
 cleaved PA are able to bind to the cell receptor. The release of
 PA20 from the remaining receptor-bound PA63 exposes the binding
 site for EF and LF, and promotes oligomerization and
 internalization of the protein.
 -1- MISCELLANEOUS: In Ref. 9 multiple mutagenesis experiments were
 performed that showed that the residues present in the small loop
 of domain 4, and not the ones in the large loop, are involved in
 receptor recognition.
 -1- SIMILARITY: BELONGS TO THE BACTERIAL BINARY TOXIN FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

Query Match 99.8% Score 3771; DB 1; Length 764;
 Best Local Similarity 99.9%; Pred. No. 1 6e-181;
 Matches 734; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 EYKOEKRLNLSSESSGCGLLGYTFESDLNFOAPVYVTSSTGDLSTPSSELEINPENOYF 60
    |||||||
DB 30 EYKOEKRLNLSSESSGCGLLGYTFESDLNFOAPVYVTSSTGDLSTPSSELEINPENOYF 89
    |||||||
OY 61 QSAIWSGFIKVKKSDYTFATSDNHNVTMWDQEVINKASNSKIRLEKGLYOIKIY 120
    |||||||
DB 90 QSAIWSGFIKVKKSDYTFATSDNHNVTMWDQEVINKASNSKIRLEKGLYOIKIY 149
    |||||||
OY 121 QRENPEKGLDFKLYMTDSQNKKEYISSDMJQPELKOKSSNSKRRKSTAGPVPDRN 180
    |||||||
DB 150 QRENPEKGLDFKLYMTDSQNKKEYISSDMJQPELKOKSSNSKRRKSTAGPVPDRN 209
    |||||||
OY 181 DGIPDSLEVEGYVDVKNRFTLSPWISNHEKKGITYYKSSPEKSTASDPYDFEKT 240
    |||||||
DB 210 DGIPDSLEVEGYVDVKNRFTLSPWISNHEKKGITYYKSSPEKSTASDPYDFEKT 269
    |||||||
OY 241 GRIDNNSPEARHPHYAAAPYIYHDMENITLSKNEQOSTONTSERTISKTSPRT 300
    |||||||
DB 270 GRIDNNSPEARHPHYAAAPYIYHDMENITLSKNEQOSTONTSERTISKTSPRT 329
    |||||||
OY 301 SEYHGNAAVHASFEDIGSGVSGNSNSTVAIDHSISLAGEKTAETMGANTADRL 360
    |||||||
DB 330 SEYHGNAAVHASFEDIGSGVSGNSNSTVAIDHSISLAGEKTAETMGANTADRL 389
    |||||||
OY 361 NANIRYVTGAPYIYVLPFTTSVLGKNOFLATIKAKENQSLQIILANNYPSKMLAPIA 420
    |||||||
DB 390 NANIRYVTGAPYIYVLPFTTSVLGKNOFLATIKAKENQSLQIILANNYPSKMLAPIA 449
    |||||||
OY 421 LNAQDFFSPTITMNYNOFLERKTKOLRLDTDOYVGNIAATYFNENGRVAVDGSNSSEV 480
    |||||||
DB 450 LNAQDFFSPTITMNYNOFLERKTKOLRLDTDOYVGNIAATYFNENGRVAVDGSNSSEV 509
    |||||||
OY 481 LPIQETARRIIFNGKDLNVERRIAANPSDELETTKPDMLKALIAFENPNGL 540
    |||||||
DB 510 LPIQETARRIIFNGKDLNVERRIAANPSDELETTKPDMLKALIAFENPNGL 569
    |||||||
OY 541 QYQCKDITEFPNFDQOTSONIKNOELNATNITYVLDKIKLAKKNNILIDRKFHVR 600
    |||||||
DB 570 QYQCKDITEFPNFDQOTSONIKNOELNATNITYVLDKIKLAKKNNILIDRKFHVR 629
    |||||||
OY 601 NNIAVGADESVEKAREVINSSTEGILLNIDKIRKILSGIYEIEDREGKKEVINDY 660
    |||||||
DB 630 NNIAVGADESVEKAREVINSSTEGILLNIDKIRKILSGIYEIEDREGKKEVINDY 689
    |||||||
OY 661 DMITNSSLROCKTEFIDFKKYNDKPLIYISNPYKKNYAAATKENTLIPSENGDSTNG 720
    |||||||
DB 690 DMITNSSLROCKTEFIDFKKYNDKPLIYISNPYKKNYAAATKENTLIPSENGDSTNG 749
    |||||||
OY 721 IKKILIFSKKGYEIG 735
    |||||||
DB 750 IKKILIFSKKGYEIG 764
    |||||||
  
```

RESULT 2
 YPBI_BACAN STANDARD; PRT; 204 AA.
 AC P13422; Q9X377;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein PX01-111.
 GN PX01-111.
 OS Bacillus anthracis.
 OC Bacterioidetes; Bacillus/Clostridium group;
 OC Bacillus/staphylococcus group; Bacillus.
 OX NCBI_TaxID=1392;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=89172073; PubMed=3148491;
 RA Melios S.L., Lowe J.R., Eden McCutchan F., Vodkin M., Leppia S.H.,
 RA Schmidt U.J.,
 RT "Sequence and analysis of the DNA encoding protective antigen of
 RT Bacillus anthracis."
 RN Genes 69:287-300(1988).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Steinere;
 RX MEDLINE=99445483; PubMed=10515943;
 RA Ohtaka R.T., Cloud K., Hampton O., Hoffmaster A.R., Hill R.K.,
 RA Keim P., Koehler T.M., Lamke G., Kumano S., Mahillon J., Menter D.,
 RA Martinez Y., Rieke D., Svensson R., Jackson P.J.,
 RT "Sequence and organization of pX01, the large Bacillus anthracis
 RT plasmid harboring the Anthrax toxin genes."
 RL J. Bacteriol. 181:6509-6515(1999).
 CC -I- SUBCELLULAR LOCATION: Membrane-bound (potential).
 CC
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M2289; AAA22636.1;
 DR EMBL: AF065404; MAD32415.1;
 DR HSSP: P13423; JACC.
 KW Hypothetical protein; Plasmid; Transmembrane.
 FT TRANSMEM 162 182
 FT CONFLICT 93 93 t->a (IN REF. 1).
 FT CONFLICT 184 204 KSCNCLLYVEVSQIMNSVEY -> NHVYILSM (IN
 FT REF. 1).
 SO SEQUENCE 204 AA; 23029 MW; E1657823AE4273FD CR=64;

Query Match 6.2% Score 235.5; DB 1; Length 204;
 Best Local Similarity 34.6%; Pred. No. 1 3e-05;
 Matches 56; Conservative 36; Mismatches 47; Indels 23; Gaps 6;

```

OY 587 MNILRDRFRHYRNINIAVGADESVEKAREVINSSTEGILLNIDKIRKILSGIYEI 646
    |||||
DB 1 MNILVRDP-YHDMNGNIYGVDDSYLKNKAKQILNMSDGVSLNLEDVQALSGYMLQI 59
    |||||
OY 647 EDTE-----GLKGYINDRYDMNLSSLRQCKTEFIDFKKYNDKPLIYISNP 693
    |||||
DB 60 KKPNSHLNPSYITILAGKDSVGBELRVLS-----DGTGTFIDFKKPEENRSLV-DPG 112
    |||||
OY 694 YVNYVYATKEN-TIINSENGDSTNGIKKILIFSKKGYEIG 734
    |||||
DB 113 DQYVYAVYKEDRNAYTRDENGNIA-NKLNLTLYLSGKIKEL 153
    |||||
  
```

RESULT 3
 MSP1_PLAFK STANDARD; PRT; 1630 AA.
 AC P04932;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
 DE (PMSA) (P190).
 GN MSP-1.
 OS Plasmodium falciparum (Isolate K1 / Thailand).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5839;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86136024; PubMed=3004972;
 RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
 RA Stuenkelberg H., Bujard H.;

"Polymorphism of the precursor for the major surface antigens of Plasmodium falciparum merozoites: studies at the genetic level.";
 EMBL J. 4:3823-3829(1985).
 [2]
 REVISIONS: SEQUENCE FROM N.A.
 RA Pan W., Tolle R., Bujard R.;
 RA Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 (Potential).
 CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
 kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 CC EMBL: X03371, CA27070.1;
 DR PIR: A25120, SAKOKI.
 DR InterPro: IPR000561, EGF-like.
 DR Pfam: PF00006, EGF_1.
 DR K1: Malattia, Merozoite, Polypeptide, Repeat, Signal, Glycoprotein;
 KM Transmembrane, GPI-anchor.
 KM SIGNAL 1 19
 FT CHAIN 20 1630 MEROZOITE SURFACE PROTEIN 1.
 FT DOMAIN 67 84 TRIPEPTIDE SG(7P) REPEAT.
 FT TRANSMEM 1614 1630
 FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 755 755 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 759 759 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 774 774 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 835 835 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 911 911 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 955 955 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1049 1049 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1156 1156 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1436 1436 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1517 1517 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT SEQUENCE 1630 AA; 187289 MW; ADBDEC3CEDA46322 CMC64;
 Query Match 4.8%; Score 179.5; DB 1; Length 1630;
 Best Local Similarity 20.4%; Pred. No. 0.13; Indels 285; Gaps 41;
 Matches 171; Conservative 111; Mismatches 272;
 OY 5 ENRL-LNSESSESSGGLGYFSDLNFOAPMVAVTSSITGDLSPSSLEINPSENOYFOSA 63
 DB 918 ENLISLGNKKNIYELIGKSSSE-NF-----YKILKDLSTPFYNE 956
 OY 64-IMGFIVKKSDEYTPATSDNHNVTMWDVDOEVINKANS-----NKIRLEKGRLY- 114
 DB 957 SFTNFVKSKAD-----INSLNDESKRKLLEDINKTKTQLSFDLYNKKYKILKERLFD 1011
 OY 115 -----QIRIOYQRENTERKGLFELYWTDSOKKKEVLSNLDLPRLAKOKSNSRKRS 168
 DB 1012 KKKTVQKKYKMOIKRLTLKEOLESKL-NSLNNPKHVL--ONFSVFPNKKKAEAEIETEN 1067
 OY 169 TSGAPVDPDRNDGIPDSLEVEGYVVDVKKRPTFLSPWISINHKKKGLRY---SSP-- 223
 DB 1068 T-----LEWTKILTKHY-----KGLVYVYNGSSPLK 1094
 OY 224 -----EKWSTASDPYSDFE-----KYTGRIKVNYSPPAR-----HPLVAAVYIYVD 265
 DB 1095 TLESEEDJEDNVALENFKVLSKLEKLDNDLNLEKKKLYLSSGHHLLA-----E 1147
 OY 266 MEMIILSNBEOGQNMDESETRTISKNTSRPHTEVYGNAGNAVHAFSDIGGSYSAFS 325
 DB 1148 LKEVYTKNK--YTGNPSSENN-----DYNNALESYKFLPECTDVAITYVS 1191
 OY 326 NSNSTVAIDHSLSLAGERTMAETMGLNADPARLNARIVNTGAPAYNNLEPT---- 381
 DB 1192 EGGSDTLQSGOPKRPASTHVGAE--NTITTSQ--NVDEVDVYLVPLGESEEDYDDL 1247
 OY 382 -ELVIGKNOGLATIKAKENOLSLAPNNVPSKMLAIALMAQNDSSSTPTMYNOPL 440
 DB 1248 GGVTVGEAVVPSYI--DNILSKI--EMEYELVILKFLA-----GVYR 1285
 OY 441 ELEKTKOLRLDQVYGNATYENGVR---RVDTGSNMSEVLPJOETARIIFNGKD 497
 DB 1286 SKTK-----QLENNVMTFNVNVKDLNRSFRKRENFNVLE-----SD 1323
 OY 498 L--NLVERRIAIVPSDPLETKPPMTLKEALIAFGNEPN--NLQYOGKDLTERDFN 553
 DB 1324 LIPYKDLTSSNVVYKDPYKFLNKKRDKPLSSVNI-----KDSIDYDIN 1368
 OY 554 PFOGT-----SONIKNOIAEL-----NATNYI-TYLDKI----- 581
 DB 1369 FANDVIGYKILSEKKSQSDLSIKKIYNDKOGENEKYLPFLNNIELTYTNDKIDFLVY 1428
 OY 582 -----KLANKNT-----LIRDRPHIDNNINAVADESVYEAAREVINS 622
 DB 1429 HLEAKVLTITYEKSNEYKELMYLKTJODKLADFKNNFVYIADLSTDTNNHNLTR 1488
 OY 623 --STEGLLINDKDI-RKILSGYI--VEIEDTBLGKEVINDRIYMLNISLRDQKFTI 676
 DB 1489 FLSTGWFEINATVTYVSLNLDGMLNSQHCYKCKOPNSCFRHLDERECKCL 1548
 OY 677 DEKAYNDKLPYLSINPKYVNVYAVTKENTIIIPSENG-----DISTINGIRKI 724
 DB 1549 NKQGDCKC--VENNPTCN-----NGCCDADAKETEDSGSNG-KKI 1590
 RESULT 4
 ID MSP1.PLAFF STANDARD. PRT; 1639 AA.
 AC P04933;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Merozoite surface protein 1 precursor (Merozoite surface antigens)
 DE (PMNSA) (P195).
 GN MSP-1.
 OS Plasmodium falciparum (Isolate Wellcome).
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OK NCBI TaxID=3648;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,
 RA Nicholls S.C., Hillman Y., Davey L.S., Tizard M., V., Schwarz R.T.,
 RA Freeman R.R.;
 RT Primary structure of the precursor to the three major surface
 RT antigens of Plasmodium falciparum merozoites.*;
 RT Nature 317:270-273(1985).
 [2]
 REVISIONS.
 RA Holder A.A.;
 RA Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 (Potential).
 CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 kDa, 42
 kDa AND 19 kDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

```

CC or send an email to license@lsb-sib.ch).
CC EMBL; X02919; CAA26676.1; -.
DR PIR; A24594; A24594.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KV Transmembrane; GPI-anchor.
FT SIGNAL 1
FT CHAIN 1
FT MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 20 1639
FT CARBOHYD 116 116
FT CARBOHYD 268 268
FT CARBOHYD 764 764
FT CARBOHYD 768 768
FT CARBOHYD 783 783
FT CARBOHYD 844 844
FT CARBOHYD 920 920
FT CARBOHYD 964 964
FT CARBOHYD 1058 1058
FT CARBOHYD 1165 1165
FT CARBOHYD 1174 1174
FT CARBOHYD 1445 1445
FT CARBOHYD 1526 1526
SQ SEQUENCE
1639 AA; 187618 MW; 2C25B6616C87F6E CRC64;

Query Match 4.8%; Score 179.5; DB 1; Length 1639;
Best Local Similarity 20.4%; Pred. No. 0.13;
Matches 171; Conservative 111; Mismatches 272; Indels 285; Gaps 41;

QY 5 ENRL-INSESSSOGGLGYFSDLNFOAPWVYTSYGDLSPSSSELENPESENOYFQSA 63
DB 927 ENLISIGKNKNIYELIGOKSSE-NF-----YKILKSDSTFEYNE 965
QY 64 IMSGFITKKEDEYTFNTSADNHVYMWVDOEVINKASNS-----NKIRLEGRIY- 114
DB 966 SFNRYVSKAD-----INSLNDESKRKLKLEEDINKTKITQLSFDLYNNKIKLERIFD 1020
QY 115 -----QIKIYOENPTEKGLDFELKYTDOSOKRKEVISNOLPLKPKKSNRKRKS 168
DB 1021 KKTYGAKKMKIKKILKLEQLESKL-NSLNPKRYVL-QNFSYFNKPKKIAETAELEN 1076
QY 169 TSNAGPIVDRNDGIPDSLEVGIVDYKKRFFLSFWATSNHKKELITY---KSSP- 223
DB 1077 T-----LENKTKILKHY-----KGLVATYINGSSPLK 1103
QY 224 -----EKWSTASDPYSDFE-----KYTGRIKIVNSPEAR-----HPLVAAPPIVHVD 265
DB 1104 TISEESIQTEENYASLENKRVLSKLEGLKMDNLKPKKLSLSSGLHHLIA-----E 1156
QY 266 MENIILSKNEDOSTONTDESETRTISKNTSTSKRTHTSEVGNAAEVHASFPIGGSVAGFS 325
DB 1157 LKEVYIKNNK---YTGNSPSENNMT-----DVNNALLESYKKRFLPEGDAVATVS 1200
QY 326 NSNSSTVAIDHSLSLAGERTMAETMGLTADTARLANANRYVNTGTPATYVNLPTT--- 381
DB 1201 ESGSDTLRQSQPKKPAFTVGAES---NTYTTTSG-VNDEVDVDYIYPIYIGSEEDYDDL 1256
QY 382 -SIVLAKNQTLATIKAKENOLSOIILAPNNYPSKKNAPILATNADDDSSPIITMYNOFL 440
DB 1257 GQVYVGEAVTSV---DNILSKI---ENEYVILYLKPLA-----GYIR 1294
QY 441 ELEKTKQALRDTDOYVGNATATYFNENGIV---RVDSGMSGEVLPJOIQTETARIIFNKO 497
DB 1295 SLAK-----QLENNVMTNNAVYKDIANSFKRREKFAVLE-----SD 1332
QY 498 L-----NLYERILAVNSDPLETTPKPTLKEALKAFGEFNPNGNIQYOGKDITFEPFN 553
DB 1333 LPIYKOLUSSWYVADYKFLKKEKRDQKFLSSVNYI-----KSDIDTIDIN 1377
QY 554 FDOQT-----SONIKNOIAEL-----NATNITY-TVIDKI-----581
DB 1378 FANDVLYGYTKILSEKYSKDLSDIKKTYINDROGENEKKYLPFLNNIETLYKTYVNDKIDLEVI 1437

```

```

QY 582 -----KLNKAKNI-----LIRKRFHYDRNNINIAVGADESVYKEAREVINS 622
DB 1438 HLEAKVLYNTEKSNVEYKIKELNYLTKTIDQKLADFKKNNFVYIADLSTDYNNHNHLL 1497
QY 623 ---STEGILLNIDDI-RKILSGYI---VEIEDEGEKKEVINDRITPLNLTSSIRDOGTPL 676
DB 1498 FLKTMVRENATVYLSNLNNGNLSQGLMNLNSHCQVKKCPQNS-KQFRLDERBECKCL 1557
QY 677 DRKKYKMDPLRYISNPNYVAVYKENTINPSENG-----DTSNGIKKI 724
DB 1558 NIKQSDKIC---VENNPYCN-----EN-----NQCCLADKCTEEDSSNG-KKI 1599

RESULT 5
ID Y67_YEAST STANDARD: PRT; 1658 AA.
AC 003661; 004988;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 187.1 kDa protein in G0A1-ERG8 intergenic region.
GN YMR219W OR YMR261.13 OR YMR959.01.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
PY SEQUENCE OF 1-711 FROM N.A.
RC STRAIN=S288C / AB972;
RA Dedman K., Brown D., Bowman S., Barrell B.G., Rajandream M.A.,
RA Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
RN [1]
RP SEQUENCE OF 608-1648 FROM N.A.
RC STRAIN=S288C / AB972;
RA Shelton J., Churcher C.M., Barrell B.G., Rajandream M.A.,
RA Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
KW Hypothetical protein.
KW
SQ SEQUENCE 1658 AA; 187137 MW; 3893F968305A757D CRC64;

Query Match 4.7%; Score 178; DB 1; Length 1658;
Best Local Similarity 18.8%; Pred. No. 0.16;
Matches 165; Conservative 139; Mismatches 286; Indels 288; Gaps 42;

QY 1 EYKQNRNLNLSSES---SSOGLGYFSDLNFOAPWVYTSYGDLSPSSSELENPESE 57
DB 290 EMBLDDLDVDSQAEKDESDQAGEHHS-VDFSKYMPRTDNTKIVIKYSDDEHYHQ 348
QY 58 QFQSAIMS-GEIKYKSD-----YFATSADN---HYTKMYDDEYVINKASNSK 105
DB 349 KISEGCAIDFQSVAYISYVDESDQAEISANAAVYHNE-VLDDDELIEDIESDS 408
QY 106 IRLERGLYQIKIYOENPTEKGLDKL-----YVTSQNKKEY 145
DB 409 -----ESQASQESQSEDDPEYKMKNEKSTSELTENTS-NSRDGFAADATYKVKY 459
QY 146 ISSDNLQQLPEL-----KQSSNSRKRKSTAGPIVDRNDGIPDSLEVE 190
DB 460 EODENDESEPKDDIIRSLDKNFKNGNNKSEYSENVLENETDAIVERNQ-IND--YB 515

```

```

QY 191 GYVYDANKRFTL---SPWISNIH- KGLITKXSS-----PEKASTASD----- 231
DB 516 GYDTGKVESDHEHS- DNLTDLAANAQOOSRNSNCPOKEQVSESLGHSNGS 573
QY 232 -----PYDPERKVTGRIDKNVSPPEARHPLVAAPLYVHWMESITLKNEDQ 277
DB 574 NLGSRSLDESEQJLPKDF---TGENNNKLTD- KGLISS- --VELEVKV---SKRK 622
QY 278 STONTSETRISKNTSTSRHTSEVHGNAEVNASPFIDGSGVSGNSNSVAIDHS 337
DB 623 LDGSTEKEVLPSTDTTIN-----NSSIGNDSITYSLDA 658
QY 338 LSLAGERTVAETMGLNTA-----DTARLNINRYVTGTAPYVNL 378
DB 659 DAISENTLOVPMEIKTTPPYEVVISESVSYSTEDNTVAMPQVEK-----TSPMND- 713
QY 379 PTTSLVLGKNOTLATIRAKENOLSGIAPNNYSPKNLAPILMANODFSPPTM- 434
DB 714 PENSL-----ND---DYKKHDLKSTLA-----LAPATKKDAFEVAGVTKSCLT 758
QY 435 -----NNQOLELEKTKOLRDPQYIGIATYFENGRRVDTGSNNSEVLPQIDETAR 490
DB 759 STSGHNTHHSKTKOVS-DLDESTEVTENETG---DENKOSKPEVANSTDK 813
QY 491 IIFNGDNLVERRIAVNSDPLETTKPDMTLEAKIAFGNPNNGNLQYOGDITEF 550
DB 814 STEDNTD---EKYFSAINTN---VTGDSGCCDIIETASVVE---NLRYCEKDMEA 862
QY 551 DFNF-DOQTSN---IKNOLA-----ELNATNYIYVLDKIKLNAMNLLJDRKFRYD 599
DB 863 EMSGCECKQKODDSKTOISFSTSPDNFOSNDYTESSTK-----YK 907
QY 600 RNNVAVGADSEVYKKAHR-EVIN-----SSTEGLLINDIDIKRI 638
DB 908 VRNSDLEDESEKELTGRAEVYDKDEESEDSEYEDYADPPGNDGSGNENYKGTAK 966
QY 639 LSGYVIEIDETGLKEVYNDRYDMLNISLQDQKTFIDFKYNDKPLTISPNRYKNV 998
DB 967 -----DTIGIYEPENKVN-----KVHEEETLEANVSSVNV 999
QY 699 YAVTKENTIIINSENGDSTNGIKKILFSSKKEIGZ 736
DB 1000 -----QKXMDHVDVINO---EAGANVENGE 1021

RESULT 6
YJL3_YEAST
ID YJL3_YEAST STANDARD: PRT: 1803 AA.
AC P47024.1887192.
DB 01-FEB-1996 (Rel. 33, Created)
DB 30-MAY-2000 (Rel. 39, Last sequence update)
DE Transposon Ty4 207.7 kDa hypothetical protein.
GN TY4B OR YJL13M OR J0780.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / FY1679;
RX MEDLINE=97103775; PubMed=8948101;
RA Czepluch C., Kordes E., Pujol A., Jauniaux J.-C.;
RT "Sequencing analysis of a 40.2 kb fragment of yeast chromosome X
RT SPT10, GCD14, RPL1, PHO86, NCA3, ASP1, CCF7, GZF3, two tRNA genes,
RT three remnant delta elements and a Ty4 transposon. ";
RL yeast 12.14471-1474(1996)
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC used by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Use by and for commercial
CC entities requires a license agreement. (See
CC or send an email to license@isb.slb.ch.)
CC -----
DB EMBL: Z49389; CA89409.1;
DB SDD: S0003649; YJL13M.
DB InterPro: IPR001584; Rve.
DB InterPro: IPR001878; Znf_CCHC.
DB Pfam: PF00665; Rve; 1.
DB SMART: SM00343; Znf_C2HC; 1.
DB Transposable element; Hypothetical protein.
SQ SEQUENCE 1803 AA; 207693 MW; 16DCD7284A8D52D3 CRC64;

Query Match 4.5%; Score 171; DB 1; Length 1803;
Best Local Similarity 19.7%; Pred. No. 0.4;
Matches 163; Conservative 123; Mismatches 273; Indels 270; Gaps 41;

QY 1 EVKQERRLNSESSESGILGYFSDLPQAPVWTSSTGDLSPSELENPENQYF 60
DB 1078 QLKKTN---HETSPKGSIG---TVNFKRTNNEISLKTGDTSLPIKLTESINHH- 1128
QY 61 QSAIWSGPIKVKASDETFPATSDNHVTWVDDQEVINKASNSKILKELGRLYQIKQY 120
DB 1129 -----SNDY-----STNKKV-----KF 1140
QY 121 QREN---PTKGLDFKLYWTDSONKKEVYSDNQLPDLQKSSNSRKRKSTSGPTVP 176
DB 1141 EKENHHPPIIDIDYDM---SDQDMESNCOGNNLKEKLYTDKNVPTDNGTNSPR- 1193
QY 177 DRDNDGIPDSLEVGTVYDVKKRFTLSPWISNHEKK-----GLTKY----- 219
DB 1194 -----LEQNTLEAGSPVQTVNKSAPFLNKEPSSILNKKFRKRHHKNNLSYSLERDKR 1247
QY 220 -----KSPPEKASTASDPYSDPEKYVGRIDKNVSPPEARHPLVAAPLYVDMENILSK 273
DB 1248 SKNRVKALIPDMETVASPKIRAIYNEAISKNPDLEKEHYKOAY---HKEIQLKDKM 1304
QY 274 NEGQONTNDSEPTTISKN---TSTSRHTSEVHGNAEVNASPFIDGSGVSGNSNSST 331
DB 1305 VFDVYKKSRSB---IPDNLVPTNTITFKRNG---YKAIYCRIDQSPPTSVIT 1358
QY 332 VALDHS---LSLAGERT-VAETMGLTADT-ARLNINRYVTGTAPYVNL-ITSL 383
DB 1359 ESLNHHIKITFLMANNNRNFMTIDINHAFLYAKLEE---IYIHHDRRC 1408
QY 384 VLKNOTLATIKAKENO-----LSOI-LAPNRYP-----SKNLAPILANODDFS 428
DB 1409 VVKLNKALIGLQSPKEMNDHLKQYLDNGIGLADNSITPOLYOTEKNNLM-IAYVDDCVI 1467
QY 429 STPTTNVNTQPL-ELEKTKOLR-----LTDQVYGNATYVNFENGAV----- 469
DB 1468 AASNEORLDEFINKLNFELKLTGTLIDVLDLD-ILQMDLVYRKVLGTLIDLLKSTIN 1526
QY 470 RVDTGSNMSEVLPQIDETARITFENGKDLNVERIAAVNP-SDPLETTKPD-----MTL 523
DB 1527 RMD--KKVYBELKIRRSSIPHSYTK-----IDPKDYLOMSEEEERGQVAKL 1573
QY 524 KEALKIAFGNPNNGNLQYOGDITEFPDNFDOOTSQNTKNOLAELNATNYIYVLDKIKL 583
DB 1574 QQLT-----GELNIV-RHKCRDYIEFAVKAARLVNYPHERFYVIKIIQYL- 1620
QY 584 NAKMNLIDKRFYDRN-----NIVAGADSEVYKKAHR-----VI-----NS 622
DB 1621 -----VRKDGIGHYDCKDKKQVIATDASVGEYDQASRGIVLWGMNIFVYSNK 1675
QY 623 STEGLLINDIDIKRIISG-----IYELERT-----EGTK----- 653
DB 1676 STNRCSVTEAEALATIEGADSETLAKYTLKEGBDNDVMTTDSKPAIGLNRSGQ 1735
QY 654 -----EYINDRYDMLNISLRQDKTFI-----DEKKY 681

```


DB 1736 PREKFTWIKTEIKIKIKIKIKITGKGNADILTKPVASADPKRF 1784

RESULT 7
SMC3_YEAST STANDARD: PRT; 1230 AA.
ID SMC3_YEAST PRT; 1230 AA.
AC P47037;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Chromosome segregation protein SMC3 (DA-box protein SMC3).
GN SMC3 OR YJ1074C OR J1049.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=M303;
RC MEDLINE=97474309; PubMed=9335333;
RA Michaelis C., Clock R., Nasmyth K.;
RT "Cohesins: chromosomal proteins that prevent premature separation of
RT sister chromatids.";
RL Cell 91:35-45(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Rose M., Koetter P., Entian K.D.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [1]
RC STRAIN=S288C;
RA Sor F.J.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN CHROMOSOME SEGREGATION IN MITOSIS. COULD BE
CC PART OF A CHROMOSOME CONDENSATION MOTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: Y14278; CAA74555.1; -
DR EMBL: Z49349; CAA89366.1; -
DR EMBL: X88831; CAA61313.1; -
DR SGD: S0003610; SMC3.
DR InterPro: IPR003439; ABC_transport.
DR InterPro: IPR003405; SMC_C.
DR InterPro: IPR003395; SMC_N.
DR Pfam: PF02483; SMC_C/1.
DR Pfam: PF02463; SMC_N/1.
DR Mitosis; ATP-binding; Coiled coil; Nuclear protein.
FT NP_BIND 32 39 ATP (POTENTIAL).
FT DOMAIN 172 482 COILED COIL (POTENTIAL).
FT DOMAIN 685 1041 COILED COIL (POTENTIAL).
SQ SEQUENCE 1230 AA; 14136 MW; B152D88b7780341f CMC64;

Query Match 4.5%; Score 169.5; DB 1; Length 1230;
Best local similarity 19.8%; Pred No. 0.28; 355; Indels 247; Gaps 38;
Matches 182; Conservative 136; Mismatches 355; Indels 247; Gaps 38;
QY 1 EYKQENRLNLSSESSGGLGYSDANFAQPAVYT--STTGGLSTPSSLEINPSENQ 58
DB 209 EMBQERKELEYNELBNKRTYQFTLVDRLNENYINQERLDGYN-----NTYSSSQ 262
QY 59 YFO-----SAIMSGFTAKVKSDEYTFATSDNHNVTWVDDQEVNKR--- 99

DB 263 YIQLDKREDMIDQVSKLLSI--EASLKIKNATDLOQAKRESEISKLTNWNVYKIDQV 321
QY 100 ---ASNSNKRILEKGLYQIK-IOYORENPTKGLDFKLYWTDSQNKKEISSDNLQPE 155
DB 322 QQIESNERQMDLDAATLKEKSIIOQRQKLSKILPVOELT-----KEAPY--KLQIAS 375
QY 156 LKQSSNSKRRKSTSAQTPVPRDNDGIPDSLEVEGYTVQVKNKRFELSPWNSIHE-KK 214
DB 376 LQQRQDILILKGGYAFKSKDERVTHSEIE-----ELKSS-----IONINLEDS 422
QY 215 GLTKYSSPEKSTASDPYSDKDEKVTGRIDKNVSPAPRHPLVAAVPIVHVDENILSKN 274
DB 423 QLOMDRTSLRQYSADID-----EEIEELIDISNGPPTKGOL-----EDFSELIHLKQK 471
QY 275 EDSQTNQNDSETRTISKNTSRHTSPVGNAPVNASPFDIGSSVSAQFSNSGYAI 334
DB 472 LSESLDTRKEIMREKQKQTVETILSDVNOR-----NVN--KSRSLAGIINKEI 525
QY 335 DHSLSLAGE-----RTMAETMG-----LNTADTALNANITYVNG 370
DB 526 TEKLISPEVSFVGLGELIKVNDKTKCAEYIGNSLPHIYVDTETATLNNELTRMG 585
QY 371 TAPIYVLPPTSLVGNQOTLATIKAKENQLSQILPNNYTPSKNLAPALAAQDDPSST 430
DB 586 GRVTF--IPLNRLISLSDVKPSPSTTQIQFTPLIKIKYEPREKA-----VKHVGKT 638
QY 431 PITMANYOFLLEKTRQLR--LDTDQV-YGNLAT--YNFNGRVVDGSMSEVLPQI 484
DB 639 IYVKDLOGKLIKKKLNALNATLDDRADRGVLTGGLDQHKRTRESLKLNNESRSH 698
QY 485 QETTARILFNCKDLNVERRIAANVS-----DPLETKPD-MTLK 524
DB 699 KILIELDFVNEELNDIDTKIDQVGNIRKVSNDRESVLTNIEVYRISLNTKKNKILE 758
QY 525 EALK-IAGFNEPKNLOFGQDITEPDFN---FPQOSNINKOLA-----E 568
DB 759 ESNATILIKLKNLTNTNFAQKLNTEFENDLOEFSSELSKEKERLESITWEISAANK 818
QY 569 LAKTN-----IYVLQIKLNKKNILIRDRPHYRNINAVG-----A 607
DB 819 LNTSALALEGTTIID--SLNLELSKILQENDLSKMSFVGAFIFGLDELKEQLQLE 876
QY 608 DESVYEAHRRVYNSST-----EELL--NIDKDIRI 638
DB 877 KESEVKQHNNAVLELGVQRETESLJAETNNKKLLEKANNQRLKLDNFQKSYEKT 936
QY 639 LSGYIVEIEDTEGLKE-----VINDRTM-----LNL-----ISSLRDQK 673
DB 937 MIRKTLVTRREELQQRIRIEGLLEPEDALVNDSDITSDOLLQKNDMTETISGLKYNK 966
QY 674 -FEIDFKRYNDR-----LPLYISNPNYKVNVAATKENTIIINPSNG 714
DB 997 RAEENKKNFERNRDLAERASELDESKDSIODLIVALKQKQVNAV---DSTFOKVSBNF 1052
QY 715 DT-----STNGIKKILIFSK 729
DB 1053 EAVFERLVPRGTATLTHRK 1072
RESULT 8
WAPA_BACSU STANDARD: PRT; 2334 AA.
ID WAPA_BACSU
AC 007833;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Wall-associated protein precursor.
GN WAPA OR N17G.
OS Bacillus subtilis.
OC Bacillales; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;

[1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE-93302506; PubMed-8316082;
 RA Foster S.J.;
 RT "Molecular analysis of three major wall-associated proteins of a
 Bacillus subtilis 168: evidence for processing of the product of a
 gene encoding a 258 kDa precursor two-domain, ligand-binding
 protein.";
 RL Mol. Microbiol. 8:299-310(1993).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168 / BGSCLAI;
 RX MEDLINE-95219088; PubMed-7704263;
 RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
 RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
 genome containing the hut and wapa loci.";
 RL Microbiology 141:337-343(1995).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168 / BGSCLAI;
 RX MEDLINE-97124196; PubMed-8969509;
 RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
 Miya Y., Fujita Y.;
 RT "Sequencing of a 65 kb region of the Bacillus subtilis genome
 containing the hic and cel loci, and creation of a 177 kb contig
 covering the gut-sacry region.";
 RL Microbiology 142:313-323(1996).
 CC -1- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
 MOTILITY, SECRETION OR DIFFERENTIATION.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
 INTO THE MEDIUM.
 CC -1- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE
 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
 C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
 MOTIF REPEATED 31 TIMES.
 CC -1- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME
 SIMILARITY TO THE REPEAT IN E. COLI RBS GROUP OF PROTEINS (RBSA-D).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
 or send an email to license@isb-slb.ch)
 CC -----
 DR EMBL: I05634; AAA22883.1;
 DR EMBL: D31856; BAA0656.1;
 DR EMBL: D29885; BAA0626.1;
 DR EMBL: D83026; BAA1168.1;
 DR EMBL: Z89124; CAB1959.1;
 DR PIR: S32920; S32920.
 DR SUBLIST: B610797; wapa.
 DR InterPro: IPR003305; CBD_6.
 DR Cell wall; Repeat; Signal; Complete proteome.
 KW Cell wall; Repeat; Signal; Complete proteome.
 FT SIGNAL 28
 FT CHAIN 1
 FT DOMAIN -29 2334 WALL-ASSOCIATED PROTEIN.
 FT REPEAT 504 869 3 X 101 AA APPROXIMATE TANDEM REPEATS.
 FT REPEAT 605 605 1-1.
 FT REPEAT 636 736 1-2.
 FT REPEAT 769 869 1-3.
 FT DOMAIN 1021 2139 31 X 21 AA APPROXIMATE TANDEM REPEATS OF
 X(4)-G-X(4)-[YF]-X-D-X(2)-G-X(4).
 FT REPEAT 1021 1040 2-1.
 FT REPEAT 1042 1061 2-2.
 FT REPEAT 1063 1082 2-3.
 FT REPEAT 1083 1102 2-4.
 FT REPEAT 1109 1128 2-5.
 FT REPEAT 1129 1148 2-6.
 FT REPEAT 1150 1169 2-7.
 FT REPEAT 1174 1193 2-8.

FT REPEAT 1199 1218 2-9.
 FT REPEAT 1219 1238 2-10.
 FT REPEAT 1240 1259 2-11.
 FT REPEAT 1260 1279 2-12.
 FT REPEAT 1280 1300 2-13.
 FT REPEAT 1300 1320 2-14.
 FT REPEAT 1320 1340 2-15.
 FT REPEAT 1340 1360 2-16.
 FT REPEAT 1360 1380 2-17.
 FT REPEAT 1380 1400 2-18.
 FT REPEAT 1400 1420 2-19.
 FT REPEAT 1420 1440 2-20.
 FT REPEAT 1440 1460 2-21.
 FT REPEAT 1460 1480 2-22.
 FT REPEAT 1480 1500 2-23.
 FT REPEAT 1500 1520 2-24 (APPROXIMATE).
 FT REPEAT 1520 1540 2-25.
 FT REPEAT 1540 1560 2-26.
 FT REPEAT 1560 1580 2-27.
 FT REPEAT 1580 1600 2-28.
 FT REPEAT 1600 1620 2-29.
 FT REPEAT 1620 1640 2-30.
 FT REPEAT 1640 1660 2-31.
 FT REPEAT 1660 1680 2-32.
 FT REPEAT 1680 1700 2-33.
 FT REPEAT 1700 1720 2-34.
 FT REPEAT 1720 1740 2-35.
 FT REPEAT 1740 1760 2-36.
 FT REPEAT 1760 1780 2-37.
 FT REPEAT 1780 1800 2-38.
 FT REPEAT 1800 1820 2-39.
 FT REPEAT 1820 1840 2-40.
 FT REPEAT 1840 1860 2-41.
 FT REPEAT 1860 1880 2-42.
 FT REPEAT 1880 1900 2-43.
 FT REPEAT 1900 1920 2-44.
 FT REPEAT 1920 1940 2-45.
 FT REPEAT 1940 1960 2-46.
 FT REPEAT 1960 1980 2-47.
 FT REPEAT 1980 2000 2-48.
 FT REPEAT 2000 2020 2-49.
 FT REPEAT 2020 2040 2-50.
 FT REPEAT 2040 2060 2-51.
 FT REPEAT 2060 2080 2-52.
 FT REPEAT 2080 2100 2-53.
 FT REPEAT 2100 2120 2-54.
 FT REPEAT 2120 2140 2-55.
 FT REPEAT 2140 2160 2-56.
 FT REPEAT 2160 2180 2-57.
 FT REPEAT 2180 2200 2-58.
 FT REPEAT 2200 2220 2-59.
 FT REPEAT 2220 2240 2-60.
 FT REPEAT 2240 2260 2-61.
 FT REPEAT 2260 2280 2-62.
 FT REPEAT 2280 2300 2-63.
 FT REPEAT 2300 2320 2-64.
 FT REPEAT 2320 2340 2-65.
 FT REPEAT 2340 2360 2-66.
 FT REPEAT 2360 2380 2-67.
 FT REPEAT 2380 2400 2-68.
 FT REPEAT 2400 2420 2-69.
 FT REPEAT 2420 2440 2-70.
 FT REPEAT 2440 2460 2-71.
 FT REPEAT 2460 2480 2-72.
 FT REPEAT 2480 2500 2-73.
 FT REPEAT 2500 2520 2-74.
 FT REPEAT 2520 2540 2-75.
 FT REPEAT 2540 2560 2-76.
 FT REPEAT 2560 2580 2-77.
 FT REPEAT 2580 2600 2-78.
 FT REPEAT 2600 2620 2-79.
 FT REPEAT 2620 2640 2-80.
 FT REPEAT 2640 2660 2-81.
 FT REPEAT 2660 2680 2-82.
 FT REPEAT 2680 2700 2-83.
 FT REPEAT 2700 2720 2-84.
 FT REPEAT 2720 2740 2-85.
 FT REPEAT 2740 2760 2-86.
 FT REPEAT 2760 2780 2-87.
 FT REPEAT 2780 2800 2-88.
 FT REPEAT 2800 2820 2-89.
 FT REPEAT 2820 2840 2-90.
 FT REPEAT 2840 2860 2-91.
 FT REPEAT 2860 2880 2-92.
 FT REPEAT 2880 2900 2-93.
 FT REPEAT 2900 2920 2-94.
 FT REPEAT 2920 2940 2-95.
 FT REPEAT 2940 2960 2-96.
 FT REPEAT 2960 2980 2-97.
 FT REPEAT 2980 3000 2-98.
 FT REPEAT 3000 3020 2-99.
 FT REPEAT 3020 3040 3-00.
 FT REPEAT 3040 3060 3-01.
 FT REPEAT 3060 3080 3-02.
 FT REPEAT 3080 3100 3-03.
 FT REPEAT 3100 3120 3-04.
 FT REPEAT 3120 3140 3-05.
 FT REPEAT 3140 3160 3-06.
 FT REPEAT 3160 3180 3-07.
 FT REPEAT 3180 3200 3-08.
 FT REPEAT 3200 3220 3-09.
 FT REPEAT 3220 3240 3-10.
 FT REPEAT 3240 3260 3-11.
 FT REPEAT 3260 3280 3-12.
 FT REPEAT 3280 3300 3-13.
 FT REPEAT 3300 3320 3-14.
 FT REPEAT 3320 3340 3-15.
 FT REPEAT 3340 3360 3-16.
 FT REPEAT 3360 3380 3-17.
 FT REPEAT 3380 3400 3-18.
 FT REPEAT 3400 3420 3-19.
 FT REPEAT 3420 3440 3-20.
 FT REPEAT 3440 3460 3-21.
 FT REPEAT 3460 3480 3-22.
 FT REPEAT 3480 3500 3-23.
 FT REPEAT 3500 3520 3-24.
 FT REPEAT 3520 3540 3-25.
 FT REPEAT 3540 3560 3-26.
 FT REPEAT 3560 3580 3-27.
 FT REPEAT 3580 3600 3-28.
 FT REPEAT 3600 3620 3-29.
 FT REPEAT 3620 3640 3-30.
 FT REPEAT 3640 3660 3-31.
 FT REPEAT 3660 3680 3-32.
 FT REPEAT 3680 3700 3-33.
 FT REPEAT 3700 3720 3-34.
 FT REPEAT 3720 3740 3-35.
 FT REPEAT 3740 3760 3-36.
 FT REPEAT 3760 3780 3-37.
 FT REPEAT 3780 3800 3-38.
 FT REPEAT 3800 3820 3-39.
 FT REPEAT 3820 3840 3-40.
 FT REPEAT 3840 3860 3-41.
 FT REPEAT 3860 3880 3-42.
 FT REPEAT 3880 3900 3-43.
 FT REPEAT 3900 3920 3-44.
 FT REPEAT 3920 3940 3-45.
 FT REPEAT 3940 3960 3-46.
 FT REPEAT 3960 3980 3-47.
 FT REPEAT 3980 4000 3-48.
 FT REPEAT 4000 4020 3-49.
 FT REPEAT 4020 4040 3-50.
 FT REPEAT 4040 4060 3-51.
 FT REPEAT 4060 4080 3-52.
 FT REPEAT 4080 4100 3-53.
 FT REPEAT 4100 4120 3-54.
 FT REPEAT 4120 4140 3-55.
 FT REPEAT 4140 4160 3-56.
 FT REPEAT 4160 4180 3-57.
 FT REPEAT 4180 4200 3-58.
 FT REPEAT 4200 4220 3-59.
 FT REPEAT 4220 4240 3-60.
 FT REPEAT 4240 4260 3-61.
 FT REPEAT 4260 4280 3-62.
 FT REPEAT 4280 4300 3-63.
 FT REPEAT 4300 4320 3-64.
 FT REPEAT 4320 4340 3-65.
 FT REPEAT 4340 4360 3-66.
 FT REPEAT 4360 4380 3-67.
 FT REPEAT 4380 4400 3-68.
 FT REPEAT 4400 4420 3-69.
 FT REPEAT 4420 4440 3-70.
 FT REPEAT 4440 4460 3-71.
 FT REPEAT 4460 4480 3-72.
 FT REPEAT 4480 4500 3-73.
 FT REPEAT 4500 4520 3-74.
 FT REPEAT 4520 4540 3-75.
 FT REPEAT 4540 4560 3-76.
 FT REPEAT 4560 4580 3-77.
 FT REPEAT 4580 4600 3-78.
 FT REPEAT 4600 4620 3-79.
 FT REPEAT 4620 4640 3-80.
 FT REPEAT 4640 4660 3-81.
 FT REPEAT 4660 4680 3-82.
 FT REPEAT 4680 4700 3-83.
 FT REPEAT 4700 4720 3-84.
 FT REPEAT 4720 4740 3-85.
 FT REPEAT 4740 4760 3-86.
 FT REPEAT 4760 4780 3-87.
 FT REPEAT 4780 4800 3-88.
 FT REPEAT 4800 4820 3-89.
 FT REPEAT 4820 4840 3-90.
 FT REPEAT 4840 4860 3-91.
 FT REPEAT 4860 4880 3-92.
 FT REPEAT 4880 4900 3-93.
 FT REPEAT 4900 4920 3-94.
 FT REPEAT 4920 4940 3-95.
 FT REPEAT 4940 4960 3-96.
 FT REPEAT 4960 4980 3-97.
 FT REPEAT 4980 5000 3-98.
 FT REPEAT 5000 5020 3-99.
 FT REPEAT 5020 5040 4-00.
 FT REPEAT 5040 5060 4-01.
 FT REPEAT 5060 5080 4-02.
 FT REPEAT 5080 5100 4-03.
 FT REPEAT 5100 5120 4-04.
 FT REPEAT 5120 5140 4-05.
 FT REPEAT 5140 5160 4-06.
 FT REPEAT 5160 5180 4-07.
 FT REPEAT 5180 5200 4-08.
 FT REPEAT 5200 5220 4-09.
 FT REPEAT 5220 5240 4-10.
 FT REPEAT 5240 5260 4-11.
 FT REPEAT 5260 5280 4-12.
 FT REPEAT 5280 5300 4-13.
 FT REPEAT 5300 5320 4-14.
 FT REPEAT 5320 5340 4-15.
 FT REPEAT 5340 5360 4-16.
 FT REPEAT 5360 5380 4-17.
 FT REPEAT 5380 5400 4-18.
 FT REPEAT 5400 5420 4-19.
 FT REPEAT 5420 5440 4-20.
 FT REPEAT 5440 5460 4-21.
 FT REPEAT 5460 5480 4-22.
 FT REPEAT 5480 5500 4-23.
 FT REPEAT 5500 5520 4-24.
 FT REPEAT 5520 5540 4-25.
 FT REPEAT 5540 5560 4-26.
 FT REPEAT 5560 5580 4-27.
 FT REPEAT 5580 5600 4-28.
 FT REPEAT 5600 5620 4-29.
 FT REPEAT 5620 5640 4-30.
 FT REPEAT 5640 5660 4-31.
 FT REPEAT 5660 5680 4-32.
 FT REPEAT 5680 5700 4-33.
 FT REPEAT 5700 5720 4-34.
 FT REPEAT 5720 5740 4-35.
 FT REPEAT 5740 5760 4-36.
 FT REPEAT 5760 5780 4-37.
 FT REPEAT 5780 5800 4-38.
 FT REPEAT 5800 5820 4-39.
 FT REPEAT 5820 5840 4-40.
 FT REPEAT 5840 5860 4-41.
 FT REPEAT 5860 5880 4-42.
 FT REPEAT 5880 5900 4-43.
 FT REPEAT 5900 5920 4-44.
 FT REPEAT 5920 5940 4-45.
 FT REPEAT 5940 5960 4-46.
 FT REPEAT 5960 5980 4-47.
 FT REPEAT 5980 6000 4-48.
 FT REPEAT 6000 6020 4-49.
 FT REPEAT 6020 6040 4-50.
 FT REPEAT 6040 6060 4-51.
 FT REPEAT 6060 6080 4-52.
 FT REPEAT 6080 6100 4-53.
 FT REPEAT 6100 6120 4-54.
 FT REPEAT 6120 6140 4-55.
 FT REPEAT 6140 6160 4-56.
 FT REPEAT 6160 6180 4-57.
 FT REPEAT 6180 6200 4-58.
 FT REPEAT 6200 6220 4-59.
 FT REPEAT 6220 6240 4-60.
 FT REPEAT 6240 6260 4-61.
 FT REPEAT 6260 6280 4-62.
 FT REPEAT 6280 6300 4-63.
 FT REPEAT 6300 6320 4-64.
 FT REPEAT 6320 6340 4-65.
 FT REPEAT 6340 6360 4-66.
 FT REPEAT 6360 6380 4-67.
 FT REPEAT 6380 6400 4-68.
 FT REPEAT 6400 6420 4-69.
 FT REPEAT 6420 6440 4-70.
 FT REPEAT 6440 6460 4-71.
 FT REPEAT 6460 6480 4-72.
 FT REPEAT 6480 6500 4-73.
 FT REPEAT 6500 6520 4-74.
 FT REPEAT 6520 6540 4-75.
 FT REPEAT 6540 6560 4-76.
 FT REPEAT 6560 6580 4-77.
 FT REPEAT 6580 6600 4-78.
 FT REPEAT 6600 6620 4-79.
 FT REPEAT 6620 6640 4-80.
 FT REPEAT 6640 6660 4-81.
 FT REPEAT 6660 6680 4-82.
 FT REPEAT 6680 6700 4-83.
 FT REPEAT 6700 6720 4-84.
 FT REPEAT 6720 6740 4-85.
 FT REPEAT 6740 6760 4-86.
 FT REPEAT 6760 6780 4-87.
 FT REPEAT 6780 6800 4-88.
 FT REPEAT 6800 6820 4-89.
 FT REPEAT 6820 6840 4-90.
 FT REPEAT 6840 6860 4-91.
 FT REPEAT 6860 6880 4-92.
 FT REPEAT 6880 6900 4-93.
 FT REPEAT 6900 6920 4-94.
 FT REPEAT 6920 6940 4-95.
 FT REPEAT 6940 6960 4-96.
 FT REPEAT 6960 6980 4-97.
 FT REPEAT 6980 7000 4-98.
 FT REPEAT 7000 7020 4-99.
 FT REPEAT 7020 7040 5-00.
 FT REPEAT 7040 7060 5-01.
 FT REPEAT 7060 7080 5-02.
 FT REPEAT 7080 7100 5-03.
 FT REPEAT 7100 7120 5-04.
 FT REPEAT 7120 7140 5-05.
 FT REPEAT 7140 7160 5-06.
 FT REPEAT 7160 7180 5-07.
 FT REPEAT 7180 7200 5-08.
 FT REPEAT 7200 7220 5-09.
 FT REPEAT 7220 7240 5-10.
 FT REPEAT 7240 7260 5-11.
 FT REPEAT 7260 7280 5-12.
 FT REPEAT 7280 7300 5-13.
 FT REPEAT 7300 7320 5-14.
 FT REPEAT 7320 7340 5-15.
 FT REPEAT 7340 7360 5-16.
 FT REPEAT 7360 7380 5-17.
 FT REPEAT 7380 7400 5-18.
 FT REPEAT 7400 7420 5-19.
 FT REPEAT 7420 7440 5-20.
 FT REPEAT 7440 7460 5-21.
 FT REPEAT 7460 7480 5-22.
 FT REPEAT 7480 7500 5-23.
 FT REPEAT 7500 7520 5-24.
 FT REPEAT 7520 7540 5-25.
 FT REPEAT 7540 7560 5-26.
 FT REPEAT 7560 7580 5-27.
 FT REPEAT 7580 7600 5-28.
 FT REPEAT 7600 7620 5-29.
 FT REPEAT 7620 7640 5-30.
 FT REPEAT 7640 7660 5-31.
 FT REPEAT 7660 7680 5-32.
 FT REPEAT 7680 7700 5-33.
 FT REPEAT 7700 7720 5-34.
 FT REPEAT 7720 7740 5-35.
 FT REPEAT 7740 7760 5-36.
 FT REPEAT 7760 7780 5-37.
 FT REPEAT 7780 7800 5-38.
 FT REPEAT 7800 7820 5-39.
 FT REPEAT 7820 7840 5-40.
 FT REPEAT 7840 7860 5-41.
 FT REPEAT 7860 7880 5-42.
 FT REPEAT 7880 7900 5-43.
 FT REPEAT 7900 7920 5-44.
 FT REPEAT 7920 7940 5-45.
 FT REPEAT 7940 7960 5-46.
 FT REPEAT 7960 7980 5-47.
 FT REPEAT 7980 8000 5-48.
 FT REPEAT 8000 8020 5-49.
 FT REPEAT 8020 8040 5-50.
 FT REPEAT 8040 8060 5-51.
 FT REPEAT 8060 8080 5-52.
 FT REPEAT 8080 8100 5-53.
 FT REPEAT 8100 8120 5-54.
 FT REPEAT 8120 8140 5-55.
 FT REPEAT 8140 8160 5-56.
 FT REPEAT 8160 8180 5-57.
 FT REPEAT 8180 8200 5-58.
 FT REPEAT 8200 8220 5-59.
 FT REPEAT 8220 8240 5-60.
 FT REPEAT 8240 8260 5-61.
 FT REPEAT 8260 8280 5-62.
 FT REPEAT 8280 8300 5-63.
 FT REPEAT 8300 8320 5-64.
 FT REPEAT 8320 8340 5-65.
 FT REPEAT 8340 8360 5-66.
 FT REPEAT 8360 8380 5-67.
 FT REPEAT 8380 8400 5-68.
 FT REPEAT 8400 8420 5-69.
 FT REPEAT 8420 8440 5-70.
 FT REPEAT 8440 8460 5-71.
 FT REPEAT 8460 8480 5-72.
 FT REPEAT 8480 8500 5-73.
 FT REPEAT 8500 8520 5-74.
 FT REPEAT 8520 8540 5-75.
 FT REPEAT 8540 8560 5-76.
 FT REPEAT 8560 8580 5-77.
 FT REPEAT 8580 8600 5-78.
 FT REPEAT 8600 8620 5-79.
 FT REPEAT 8620 8640 5-80.
 FT REPEAT 8640 8660 5-81.
 FT REPEAT 8660 8680 5-82.
 FT REPEAT 8680 8700 5-83.
 FT REPEAT 8700 8720 5-84.
 FT REPEAT 8720 8740 5-85.
 FT REPEAT 8740 8760 5-86.
 FT REPEAT 8760 8780 5-87.
 FT REPEAT 8780 8800 5-88.
 FT REPEAT 8800 8820 5-89.
 FT REPEAT 8820 8840 5-90.
 FT REPEAT 8840 8860 5-91.
 FT REPEAT 8860 8880 5-92.
 FT REPEAT 8880 8900 5-93.
 FT REPEAT 8900 8920 5-94.
 FT REPEAT 8920 8940 5-95.
 FT REPEAT 8940 8960 5-96.
 FT REPEAT 8960 8980 5-97.
 FT REPEAT 8980 9000 5-98.
 FT REPEAT 9000 9020 5-99.
 FT REPEAT 9020 9040 6-00.
 FT REPEAT 9040 9060 6-01.
 FT REPEAT 9060 9080 6-02.
 FT REPEAT 9080 9100 6-03.
 FT REPEAT 9100 9120 6-04.
 FT REPEAT 9120 9140 6-05.
 FT REPEAT 9140 9160 6-06.
 FT REPEAT 9160 9180 6-07.
 FT REPEAT 9180 9200 6-08.
 FT REPEAT 9200 9220 6-09.
 FT REPEAT 9220 9240 6-10.
 FT REPEAT 9240 9260 6-11.
 FT REPEAT 9260 9280 6-12.
 FT REPEAT 9280 9300 6-13.
 FT REPEAT 9300 9320 6-14.
 FT REPEAT 9320 9340 6-15.
 FT REPEAT 9340 9360 6-16.
 FT REPEAT 9360 9380 6-17.
 FT REPEAT 9380 9400 6-18.
 FT REPEAT 9400 9420 6-19.
 FT REPEAT 9420 9440 6-20.
 FT REPEAT 9440 9460 6-21.
 FT REPEAT 9460 9480 6-22.
 FT REPEAT 9480 9500 6-23.
 FT REPEAT 9500 9520 6-24.
 FT REPEAT 9520 9540 6-25.
 FT REPEAT 9540 9560 6-26.
 FT REPEAT 9560 9580 6-27.
 FT REPEAT 9580 9600 6-28.
 FT REPEAT 9600 9620 6-29.
 FT REPEAT 9620 9640 6-30.
 FT REPEAT 9640 9660 6-31.
 FT REPEAT 9660 9680 6-32.
 FT REPEAT 9680 9700 6-33.
 FT REPEAT 9700 9720 6-34.
 FT REPEAT 9720 9740 6-35.
 FT REPEAT 9740 9760 6-36.
 FT REPEAT 9760 9780 6-37.
 FT REPEAT 9780 9800 6-38.
 FT REPEAT 9800 9820 6-39.
 FT REPEAT 9820 9840 6-40.
 FT REPEAT 9840 9860 6-41.
 FT REPEAT 9860 9880 6-42.
 FT REPEAT 9880 9900 6-43.
 FT REPEAT 9900 9920 6-44.
 FT REPEAT 9920 9940 6-45.
 FT REPEAT 9940 9960 6-46.
 FT REPEAT 9960 9980 6-47.
 FT REPEAT 9980 10000 6-48.
 FT REPEAT 10000 10020 6-49.
 FT REPEAT 10020 10040 6-50.
 FT REPEAT 10040 10060 6-51.
 FT REPEAT 10060 10080 6-52.
 FT REPEAT 10080 10100 6-53.
 FT REPEAT 10100 10120 6-54.
 FT REPEAT 10120 10140 6-55.
 FT REPEAT 10140 10160 6-56.
 FT REPEAT 10160 10180 6-57.
 FT REPEAT 10180 10200 6-58.
 FT REPEAT 10200 10220 6-59.
 FT REPEAT 10220 10240 6-60.
 FT REPEAT 10240 10260 6-61.
 FT REPEAT 10260 10280 6-62.
 FT REPEAT 10280 10300 6-63.
 FT REPEAT 10300 10320 6-64.
 FT REPEAT 10320 10340 6-65.
 FT REPEAT 10340 10360 6-66.
 FT REPEAT 10360 10380 6-67.
 FT REPEAT 10380 10400 6-68.
 FT REPEAT 10400 10420 6-69.
 FT REPEAT 10420 10440 6-70.
 FT REPEAT 1044

DT 01-OCT-1996 (Rel. 34, last sequence update)
 DT 01-OCT-1996 (Rel. 34, last annotation update)
 DB Hypothetical 222.8 kDa protein ClpF.06c in chromosome I.
 GN SPAC1F3.06c.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 RX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RL Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: 270690; CA94624.1;
 DR Hypothetical protein.
 DR SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;
 SQ
 Query Match 4.38; Score 162; DB 1; Length 1957;
 Best Local Similarity 20.4%; Pred. No. 1.3;
 Matches 177; Conservative 132; Mismatches 328; Indels 232; Gaps 40;
 OY 2 VKQENLLNESE-----SSOGLGYTFSDLNFO--APMYVTSSTTGDLSPSS 49
 DB 129 VTQKSNLNEKQVRSKLALHEHNGILSLQSLSSKKMKNTSVTLTSEEDVSYFQKK 188
 OY 50 LENISENQYQSAIWS--GFIKY-----KKSDEYTFATSDNHVVMVMDQYINRAS 101
 DB 189 LVTMESENSAKOSAVYDLSROLLVTYEKLDKKEDEYKEDVSAIKASIAEQAQNSNL 248
 OY 102 NSKRIELEK--GRLYQIKIYQRENPT--KGLDFKLYMTDSQNKKEVSSDNQJPE 155
 DB 249 RGRQERLEKLVYSNKTIVSTLQTFENSIRACVTLQEKL-----EKCAINEDSLLEE 302
 OY 156 LKQKSN-----SRKKRSTAGPTVPDRNDGIPDLSLEVGYVDYVKNKR----- 200
 DB 303 LKHNVAANTSDAIVHKKLTIEDLSTRISEFN-----LKSEHDTLSIKNEKLEKLNMTI 356
 OY 201 -TFLSPMSTNHKKGLTKYKSS--PEKMTASDPYSDPEK-----VTGRD----- 244
 DB 357 GSIKDSRTNSQLEELNVELKESNPTIHSQITDSEKLSLSPQENKSLKGSIDERYNNLS 416
 OY 245 -----KWS--PEARHPLVA-----YPIVHDM----- 266
 DB 417 SKDKMYKQVSSOLEEARSLSLAHTGKLAELINSEKDFONKRIKDEFEKIDRACLNSSN 476
 OY 267 ---ENILSKNEDOSTOND--SETRTISKNTS-----RTTSEVHNAEVAHAFPD 315
 DB 477 ELKEKSLIDKRDDELNMLREQIKQKKVSESSQSLQSLQIDILINKKHEVYSQSLNE 536
 OY 316 IGGSVAGSAGSNSNS--STVAIDHSIAGERTVAETMAGTADTAPRANANIRVNTG 370
 DB 537 LKGELOTEITNSSEHLSQSLTLAKKEAVALVNNELSEK--NSIQTL-CNA----- 585
 OY 371 TATIVAVLPTTSLVIGNO--TLATIKAKENQSLIAPANNYPKRIAPALNAODD 426
 DB 586 ---FOKRLAKSVQMLKENQNSISLDTSPKLMESHOELENNHQTIRKL-----KD 634
 OY 427 FSGSTPTIMNYQFLEKTKQRLDTPDYGNATYENNGVAVYDVGSMKSVLPQOE 486
 DB 635 TSS-----KQOQLQERANFQKESSTISDENMDLTK-----LKLIEE 672
 OY 487 TTRATIFNGKDLNIVERKIAAANVPDPLETTKKPDM-TLKEALKI-----AFGENEPMQ 541

DB 673 SNKSLIKQOEVLSTKNTI-----QTLKEDLKSEELRFSKLEKLNREYIDNL- 722
 OY 542 YQKQDIT-EFDENFDQOTSONIKRQALAEALNATYITVLDKIKLNKNTILIDKRFHYDR 600
 DB 723 -KQKHETLEAQRDLNLSLSDAKNTNMLISSELTSSDVRKRLVANVETLTQDSK----- 776
 OY 601 NNIAVGADEVYKAEAREVINSSTEGILLNIDKIR-----KILSGYIVEI-----DT 649
 DB 777 -----AMKSFSLVNSYOS--ISNLYHELIDHDHVMQSQNNLTLESSEKLTDC 824
 OY 650 EGIKE---VINDRDMINISLSDPGCTFIDFKYNDKLPYISNPYVANYAVATEMT 706
 DB 825 ENLTQNNLTLDVQKLMKHVNOESKV-SELKEVNGSLSDLKLN-RSSLVN-AISNDQ 882
 OY 707 IINP-----SENGPT-----STNGIKI 724
 DB 883 ILTQLAELSKNYDSLEQESQNLGSLSL 911
 RESULT 11
 USOL YEAST STANDARD; PRT; 1790 AA.
 AC P25386;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Intracellular protein transport protein USOL.
 GN USOL OR IMT1 OR YD058W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 RX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=X2180-1A;
 RC MEDLINE=91185402; PubMed=2010462;
 RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
 RA Yamaaki M.;
 RT "A cytoskeleton-related gene, usol, is required for intracellular
 RT protein transport in Saccharomyces cerevisiae."
 RL J. Cell Biol. 113:245-260(1991).
 RN [2]
 RP SEQUENCE OF 782-1790 FROM N.A.
 RA Hostetter M.K., Herman D.J., Bendel C.M., McCellan M., Tao N.,
 RA Kennedy K.E.;
 RL Submitted (Feb-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-8 FROM N.A.
 RA Bai Y., Symington L.S.;
 RL Submitted (May-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
 CC COMPLEX.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
 CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
 CC ER AND THE GOLGI COMPLEX.
 CC -1- DOMAIN: THE GOLGI TAIL. SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
 CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
 CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
 CC -1- SIMILARITY: BELONGS TO THE VDP/USOL/YBLO47C FAMILY.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X54378; CA93253.1;
 DR EMBL: U03188; AAB00143.1;
 DR EMBL: U03688; AAB06659.1;
 DR PIR: A38455; A38455.

DR HSP: P80220; IDP:
 DR SCD: S0002216; US01.
 DR Interpro: IP8002017; Spectrin.
 KW Transprot: Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
 FT DOMAIN 1 724 GLOBULAR HEAD.
 FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
 FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
 FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
 FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
 FT CONFILIC 847 847 G -> E (IN REF. 2).
 FT CONFILIC 924 924 E -> K (IN REF. 2).
 FT CONFILIC 1253 1253 V -> I (IN REF. 2).
 FT CONFILIC 1319 1319 I -> V (IN REF. 2).
 FT CONFILIC 1461 1461 N -> S (IN REF. 2).
 FT CONFILIC 1581 1581 G -> S (IN REF. 2).
 FT CONFILIC 1600 1600 I -> V (IN REF. 2).
 FT CONFILIC 1661 1661 R -> S (IN REF. 2).
 FT CONFILIC 1772 1772 D -> DEEDDE (IN REF. 2).
 FT SOURCE 1790 AA; 206424 MW; 6C2B216E9FD4818 CRC64;

Query Match 4.3%; Score 161; DB 1; Length 1790;
 Best Local Similarity 18.5%; Pred. No. 1.2;

Matches 144; Conservative 151; Mismatches 263; Indels 220; Gaps 37;

QY 1 EVKQENRLNLSSESSQGLGYFSDNFOAPMWTSTGDD-----LSIPSELE- 51
 DB 1036 QYKEE--ITSKSDSKD---EYESQISLTKERLTATANDENVNKSISELTKRELEA 1089
 QY 52 -----NIPSENOYFQSAIWSGFIKVKSDY-----TFATSAHNMVMDQEV 96
 DB 1090 ELAAVKNKLNLETKLETKSEKAKVEVNEHLEKEKIQLEKATETKQQLSLKANLES 1149
 QY 97 INKANSNKTLEKGRLOYKIR-----QY-----OREPREKLDKRLW- 136
 DB 1150 LKEHEHDLAQLK---YEOJANKEROYBELSOLNDEITSTQENESIKKNDLEGE 1206
 QY 137 -----TSONKREYISSDNLOPELKQKSNRRKSTSGPYVDRNDQIPSL 187
 DB 1207 VKKAMSTSEBONSKESEIDALNDIKELKKN-----ETENASLESI 1250
 QY 188 E-VEGYIVDV--NKRTELSPWISNHEKKGFLTKYKSPKSTADSPDYSPKVTGR 243
 DB 1251 KVSSEYVRIKELDDECNREKREVESELD---KLKASEKNSKYLELOKESKEIKEL 1305
 QY 244 DKVNSPEARHPLVAAPYIVHVDENII-LSKNEDOSTONTDETETRTISKNTSTRTSE 302
 DB 1306 DAKTTE-----LKIOLEKITLSKAKES---ESELRLKTKTSSEERKNAE 1349
 QY 303 ---VHGADEVASFPDIGSVAGFSNSNSTVAIDHSLSLAGERTWETGLTADTA 358
 DB 1350 QLEKIKNEIOIKNAPE---KERRLNBSSTITQEVY-----INTLE- 1391
 QY 359 RLNANIRVYNGTAPYIVNLPPTSIVLKNQTLATIKAE-----NOLSOILAPNNYPS 413
 DB 1392 -----DELIRJONEN--ELKAKELIDNRSLEKVSISLDELLE 1427
 QY 414 KNLAPIALNADDPSS--PTPMYNOFLEKTKURLDIDVOYGIAIYVNGRGRV 471
 DB 1428 EKQNTIK-SLQDELSTYKDKITRNDKELSLERDNKRLDSEK-----EQLRAQ 1476
 QY 472 DTSMSSEVLOIOTETAR--IINGDL---NLVGRILAVNSDPLET-TKRPMT 522
 DB 1477 ESKAKVEBLKLEBSSKEKELKSKEMAKKLESTISINTELAKS--METIKRSPK 1534
 QY 523 LKALAKIAPGNPNGLYOQKOTTEPDNPOOTSONIKNL-----ALNAT--N 573
 DB 1535 LEOKSKA---EDYIKNLQHEKSDLS-RINSEKQIEKLSKRTIANSGLSELTVAOE 1590
 QY 574 IYTVLQIKLAKAKNILLDKRFHYDRNNIAVGADESVYKHAHEVINSSTEGILLNDK 633
 DB 1591 LNNAOEKIRINABENTYLSK-----LEDIERELDKQAE---IKNSO 1630

QY 634 DIRKILSGYVEID-----TEGIKEYINDRDMINISLROKSTFIDEKRYND 683
 DB 1631 EKEKILTSRUKLEIDELDSTOOKAKNSEERRAEVRKFOVEKSQLKEMKLETKRYND 1688
 RESULT 12
 ID BND.CLOBO STANDARD: PRT; 1276 AA.
 AC P19321;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Botulinum neurotoxin type D precursor (EC 3.4.24.63) (BoNT/D)
 CN BoNTD
 OS Clostridium botulinum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 CC Clostridium.
 CC NCBI_TaxID=1491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN-BVD/-3;
 RX MEDLINE=91016853; PubMed=2216736;
 RA Binz T., Kurazono H., Popoff M.R., Eklund M.W., Satauchi G.,
 RA Kozaki S., Kriesglstein K., Henschen A., Gill D.M., Niemann H.;
 RT "Nucleotide sequence of the gene encoding Clostridium botulinum
 RT neurotoxin type D.";
 RL Nucleic Acids Res. 18:5556-5556(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RP STRAIN-CB16;
 RX MEDLINE=93042276; PubMed=1420572;
 RA Satauchi G., Ohya T., Watanabe T., Inoue K.;
 RT "The complete amino acid sequence of the Clostridium botulinum type D
 RT neurotoxin, deduced by nucleotide sequence analysis of the encoding
 RT phase d-16 phi genome.";
 RL J. Vet. Med. Sci. 54:905-913(1992).
 RN [3]
 RP SERIAL SEQUENCE.
 RP STRAIN-BDSN, AND D-1873;
 RX MEDLINE=89359741; PubMed=2668193;
 RA Koriishi K., Syuto B., Kubo S., Oguma K.;
 RT "Molecular diversity of neurotoxins from Clostridium botulinum type D
 RT strains.";
 RL Infect. Immun. 57:2886-2891(1989).
 RN [4]
 RP IDENTIFICATION OF SUBSTRATE.
 RP MEDLINE=94230352; PubMed=8175689;
 RA Yamasaki S., Baumeister A., Binz T., Blas J., Link E., Cornille F.,
 RA Roques B., Eykse E.M., Suedhof T.C., Jahn R., Niemann H.;
 RT "Cleavage of members of the synaptobrevin/VAMP family by types D and
 RT F botulinum neurotoxins and tetanus toxin.";
 RL J. Biol. Chem. 269:12764-12772(1994).
 CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
 CC ENDOPEPTIDASE THAT CLEAVES THE 60-KDa-1-LBU-OL BOND OF
 CC SYNAPTOBREVIN-1 AND -2.
 CC -2- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
 CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. NO
 CC detected action on small molecule substrates.
 CC -3- SUBUNIT: DISULFIDE-LINKED HOMODIMER OF A LIGHT CHAIN (L) AND A
 CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
 CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIANTE CHANNEL
 CC FORMATION AND TOXIN BINDING, RESPECTIVELY.
 CC -4- SUBCELLULAR LOCATION: Secreted.
 CC -5- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
 CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
 CC -6- MISCELLANEOUS: BOTULINUM TYPE D NEUROTOXIN IS SYNTHESIZED BY D
 CC STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE
 CC BACTERIOPHAGE.

[illegible]

```

Db      688 NCLEBRVRCWKOSYOMVAVSNMLSRITTOFNHNINOMYDS-----LSYQAOIAKAKIDL 740
Oy      435 NTCNLELEER-----TKOLRDTQOYVGN-----ATVFNCGARVVDGNSN 476
Db      741 EKKYKSSDDEKNTKSOVENLKNLSADVKISEAMNNNNNNNFRFRECSTVYDFKR-----790
Oy      477 WSEVYPOLOETARIIIFNGKD--LNL-----YERLAAVNSDPLETFKPMKLT 523
Db      791 ---MLPQVDELKMPDLKFTKTELTIELINDSHNIIIVGEVIRLAKVKNES--FENMP-----841
Oy      524 KEALKAIFG--FENPQNLQYOGDIEPEFNDQGS-----ONKNOLAELATNATY--YTV 577
Db      842 ---FNIFSTNLSL---KDIINEFNSIDKSLISLONKKNLVDTSYAAEVY 891
Oy      578 LKRLKAKNNKILLIDKRFHIDRRNNIAGADESVYKAEHREVINSREGILLINDIKRK 637
Db      892 GNVQVLN---TITNDFKTLSSGKTLIVANNLNLTSA---IYENSSVSWIKISKDLTN 945
Oy      638 IISGYV--ELEDTEGLEVINDRYMDMLNISLSLQDKGFIDEKYNKLIPLY---ISNP 692
Db      946 SHNETTIINSIDENSOMKLCIRNG---NIEMILQVY-----KRYKSLIFDYSLSHT 996
Oy      693 NTKVMVYAVTKENTII 708
Db      997 GYTNKMFVTVITNNIM 1012

RESULT 13
PATL_SCHPO          STANDARD;          PRG: 1385 AA.
AC      013735; G9UTR0;
DF      15-JUL-1998 (Rel. 36, Created)
DF      16-OCT-2001 (Rel. 40, Last sequence update)
DF      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Actin interacting protein 3 homolog.
PATL OR SPAC15A10.15 OR SPAC15L.01.
OS      Schizosaccharomyces pombe (Fission yeast).
OC      Eukaryota; Fungi; Ascomycota; Schizosacchariomycetes;
OC      Schizosaccharomycetales; Schizosaccharomycetaceae;
OC      Schizosaccharomyces.
OC      NCBI_TaxID=4896;

[1]
RN      RP
RN      RP SEQUENCE OF 1-1033 FROM N.A.
RN      RA STRAIN-972;
RN      RA Murphy L., Harris D., Wood V., Barrell B.G., Rajandream M.A.;
RN      RA Submitted (JUN-1997) to the EMBL/Genbank/DBJ databases.
RN      RN [2]
RN      RP SEQUENCE OF 1023-1385 FROM N.A.
RN      RA STRAIN-972;
RN      RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Lyne M.H.;
RN      RA Submitted (AUG-1999) to the EMBL/Genbank/DBJ databases.
RN      RN [3]
RN      RP IDENTIFICATION, AND GENE NAME.
RN      RX MEDLINE-20143585; Pubmed-10679021;
RN      RX Jin H., Amberg D.C.;
RN      RA "the secretory pathway mediates localization of the cell polarity
RN      FT regulator Alp3p/Bud3p.";
RN      RL Mol. Biol. Cell 11:647-661(2000).
RN      RL
RN      RL -1- FUNCTION: INVOLVED IN THE ORGANIZATION AND/OR FUNCTION OF THE
RN      ACIN CYTOSKELETON
RN      CC -1- SIMILARITY: TO YEAST BUD6.

-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC      or send an email to license@isb-sdb.ch).
CC      EMBL; 297208; CAB10112.1;
CC      EMBL; AL109770; CAB52420.1;
DR

```



```

OY 167 R-----STAGPTVPHDNDGIPDSLEVEGYVDANKRTEFLSP-----WISNIHEKG 215
DB 248 RRRGSSSGNDNDNNDDDDANEDDSANL-----TKATIISEFLDIWLIS-----ES 299
OY 216 LTKYKSPK WSTASDPISDEKYTGKIDKNVSPARAPVLAAPVYVDMENITLSKN 274
DB 300 LVNOSTYLNKIMLSITNOP-----NENSS-NSPLV-----PLEFKINONILTR- 341
OY 275 EDQ-----STONTDSEF-----RTISKNT 293
DB 342 QDDYLNPIFERSEFVDMKLVHVDISLMDPEFKITSDTSEPGIIEIYVDMONISKL 401
OY 294 STSRITSEVGNAEVHASFIDIGSVAGFSNSNSVVAI-DHSLS--LAGERTWETM 350
DB 402 SPFNKESP-----ADIOACVGFALALAIASANAPLDDISIGPNSITROJASPRESIATLV 457
OY 351 GLVWADTALNANIRVNGTAPVYNVLPSTSLVGLKNQTLATIKAKENOLSOILAPNXY 410
DB 458 DI-----MINORBALNT-----TVSYVI--ELIRKNSQYQVNLITTTIKT 498
OY 411 YPSKNIAPIAL-----NADDFSSPTITMNTNPFLEETKQOLQDLYOYGNITVY 462
DB 499 HPSNDPITLIGYILTKRFSNHLSD-----FQIITIDENDANILPHENOLHEKRPYL 550
OY 463 NFENGVRVDTGSMSEVLPQIOETTAIIFNGKDLNVERIAVANP--SDPLETKPD 520
DB 551 GFERFV-----VELIAELLHCSNNGMLNSKRERARARRRORVRSOLSHLDALND 602
OY 521 MTL--REALKIAFG-----FNEPNCNLOYGKDI-----TEPDFNF--QOTSQ 560
DB 603 LSTEEKQOLKTRHSPRTDHDKNNKIDNDNDNDESDYDEDESEFELPIYNNKQ 662
OY 561 NIKQALAEINNTNITVYLDKIRLNKAKNIIIRKRFHYDRNNIAVAGD-----ESVVEA 615
DB 663 NIKLRDPVGT-----YKSNMYMRD--CFQNNELFLTPNMMFWHVAVIDI 708
OY 616 HREYIN-----STEGGLINID--KOIKIISGYIEI--DTEGLKEVINDRDMNI 665
DB 709 IQOIFNGRMDSYNSPLVSLFLKLSQYFWTDIYVSEKGDVRSRSPYIDBPNEFKI 768
OY 666 SS---LR--OGKFTFIDFKYVN-----DKLPLYISN-----691
DB 769 TTDFILRGQDSYFYELKRMNLGTMGHVILAEVYVRSKIYVDIISYNSNLOTDEM 828
OY 692 PNKVVVYVTK-----ENTINPSENGDTSTN 719
DB 829 QYISEELNEMTMMYSKILGGSYIDCGNIIIPOLDPNTIVLP--NGDASN 880

RESULT 15
CAGA_HELPJ STANDARD; PRT; 1167 AA.
ID CAGA_HELPJ 09GLT1;
AC 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytotoxicity associated immunodominant antigen (120 kDa protein)
DE (CAG pathogenecity island protein 26).
DE CAGA OR CAI OR CAG26 OR JHP0495.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moll D.T., King B.L., Brown E.D., Dolg P.C.,
RA Smith D.R., Noonan B., Guild B.C., deGange B.L., Carmel G.,
RA Tummino P.J., Caruso A., Ulla-Nickelsen M., Mills D.H., Ives C.,
RA Gibson R., Weiberg D., Mills S.D., Jiang Q., Taylor D.E., Voyis G.F.,
RA Trust J.J.;
RT Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.;

```

```

RL Nature 397:176-180(1999).
CC -! FUNCTION: MAY BE NECESSARY FOR THE TRANSCRIPTION, FOLDING, EXPORT,
CC OR FUNCTION OF THE CYTOTOXIN.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AE001483; AAD06073.1;
CC InterPro: IPR004355; IYSEC_CAGA.
CC PRINTS: PRO1553; TYPE4SSCAGA.
CC Antigen: Complete proteome.
CC FT DOMAIN 246 249 POLY-THR.
CC FT DOMAIN 882 889 POLY-ASN.
CC SQ SEQUENCE 1167 AA; 129729 MW; FD5E86B81CEBD0F2 CRC64;

Query Match 4.18; Score 156.5; DB 1; Length 1167;
Best Local Similarity 20.68; Pred. No. 1.2;
Matches 168; Conservative 136; Mismatches 29%; Indels 213; Gaps 45;

OY 3 KOENELNSESSESSGGL--GYFSDL-NF-----QAPVYVSTFGDLSIPSELE- 51
DB 480 KADKALREKATTLGSLKHGVDVYVSNFYTNASKSPDKGVGATNG--VSHLTA 534
OY 52 -----NIPSENOYFQSAIMSGFIRVKKSDYEYFATNSADNIV--WNVDD--QEVIN 98
DB 535 GFSKVAVFNPNLNNIAITV---VRODLEDKLAKGLSPOEANKLYVDELSSNKELVQ 590
OY 99 KASNSNKRLE--KORLYQIKIOYQENPTEKGLFEKLYWTDONKKEVISDNLOLPE 155
DB 591 KALNFKNAVNAEAKNTGNDYEVK--QKQKDLKSL-----KKRRELEKVAK--N 635
OY 156 LKQSNRSRKRKSTSGPTVDR-----DNDGIPDSLEVEGYVYVHKKRFRFLSPVTSNT 210
DB 636 LESKGNKNMKEAKSQANSQKDELFALINKKANRDAI--AYON--CKIKRELSDKLENT 694
OY 211 HKKGLTKYKSPKSTASDPYSDEK-----YNGRT--DKNS--SPEARHPVAAPIV 262
DB 695 N--KDLADPFSKDEPKNGKN--KDFSKAETILAKGYSKVDGILNPEVIS-----741
OY 263 HYDMENIILSKNDOSTQTDSETRTYSKATSTSTHTSEYHGM/EVHASFIDIGSVSA 322
DB 742 --KVENLMAALNEFKNGKNKDFSKVTOAK--SDLENSIKDVIIN--IDKYVDNLQNAVY 797
OY 323 GFSNSNSTVAIDHSLSLAGERTWETMGLNTADTARLANIRVNG--TAPIYVNLPTT 381
DB 798 AKATGDFGV-----EOLALADIKNFSKEOLAQDQKQKNDI--NTGKNSALVQ-----842
OY 382 SLVGLKQNOTATIKAKENOLSOILAPNPNYPSKNIAPL--ALNAQ--DDF-----S 428
DB 843 SVKNVNGNTVY-----GNGLSKAEATY--LSKNFSQDKKELNAKLGNNNNNNNGLENS 894
OY 429 SPPTIMYVNFLEETKQOLRLTDQOYGVNATYFENGARVDTGSMSEVLPQIOETT 488
DB 895 TEPV--YVQAVKVAKIDRL--DQIASGLG-----DVGOAASFILKRHDV-- 937
OY 489 ARIINCKDLNVERIAVANPSPLETKPPM--TIKEALK--AFGPNPNCNLOYOG 544
DB 938 -----DLS-----KVLGSANHEPIYATIIDLGSPFLKRHDV-----971
OY 545 KITEPEFNPDOOTSONIKN--QIAELNNT--NIYVLDKIKLANMNIILDKRFRH 598
DB 972 DDLSSVGLSSEKQTLQKIDNINQNAVSAKASHFDNDQIDKLSKSTKNV-----1023
OY 599 DNNNIANGADESVKAEHREYINSDEG--LLNLT--DKQIDRGI NGIYVEIDETGLAE 654
DB 1024 ---NLYVESAKKVPYSLSKLDNIAVNSHTRINSVKNKGTINKATGMLTO--KNSEWLK- 1078

```


OY 655 VINDRDMANISLRQDCKTEIDF-----KKYND-KLPLYISNP-----NKYVW 698
Db 1079 LVNDKIVAHNVGSAPLSAYDKIGFNOKNKOYSDSEFSTRLSNAVKDIKSGFVQFLTNI 1138
OY 699 YAVTKENTLIINSENGDSTNGIKILIFSKKGYE 733
Db 1139 FSMGYSYLMKASVBEHVKNNTN-----TKGFO 1165

Search completed: October 4, 2002, 14:39:21
Job time: 271 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 14:35:25 ; Search time 34.15 Seconds

(without alignments) 3728.383 Million cell updates/sec

Title: US-09-848-909-21

Sequence: 1 EVKQENRLNLSSESSQGL.....STNGIKLILPSKYEIGZ 736

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP archaea: *
2: SP bacteria: *
3: SP fungi: *
4: SP human: *
5: SP invertebrate: *
6: SP mammal: *
7: SP plant: *
8: SP organelle: *
9: SP phage: *
10: SP plant: *
11: SP rodent: *
12: SP virus: *
13: SP vertebrate: *
14: SP unclassified: *
15: SP virus: *
16: SP bacteriophage: *
17: SP archaea: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3702	98.0	741	2	Q937W2 bacillus an
2	3702	98.0	743	2	Q937W3 bacillus an
3	890.5	23.6	876	2	Q32739 clostridium
4	890.5	23.6	879	2	Q06498 clostridium
5	888.5	23.5	876	2	Q9KH41 clostridium
6	884	23.4	875	2	Q46221 clostridium
7	826	21.9	721	2	Q46171 clostridium
8	200.5	5.3	2276	2	Q93716 staphylococ
9	199	5.3	4688	16	Q9PQ08 ureaplasma
10	188	5.0	2647	5	Q9U4X0 plasmodium
11	186.5	4.9	2771	5	Q26216 plasmodium
12	186	4.9	1227	16	Q97K41 clostridium
13	183.5	4.9	2529	16	Q25579 helicobacte
14	183	4.8	2747	5	Q9BJX9 plasmodium
15	183	4.8	4152	2	Q9ZHL3 haemophilus
16	182.5	4.8	1125	16	Q98PM9 mycoplasma

17	181	4.8	2269	5	Q26223 plasmodium
18	180.5	4.8	149	2	Q9RM77 clostridium
19	178.5	4.7	149	2	Q9RM78 clostridium
20	178.5	4.7	1072	16	Q9CF64 lactococcus
21	178.5	4.7	1387	5	Q9GZ76 plasmodium
22	178	4.7	752	16	Q98P19 mycoplasma
23	178	4.7	1302	2	Q49547 mycoplasma
24	178	4.7	2178	2	Q46149 clostridium
25	178	4.7	4019	2	Q9ZHL0 haemophilus
26	178	4.7	6713	16	Q93186 staphylococ
27	178	4.7	6713	16	Q93186 staphylococ
28	176.5	4.7	3254	2	Q9RM80 clostridium
29	176.5	4.7	3254	2	Q9RM80 clostridium
30	176	4.7	1193	2	Q45914 clostridium
31	174.5	4.6	3130	5	Q9RM79 plasmodium
32	174.5	4.6	3130	5	Q9RM79 plasmodium
33	173.5	4.6	821	3	Q08581 saccharomyc
34	173.5	4.6	1193	2	P71107 clostridium
35	172	4.6	1939	5	Q25662 plasmodium
36	171.5	4.5	1365	2	Q49525 plasmodium
37	171.5	4.5	1365	2	Q49525 plasmodium
38	171.5	4.5	2399	16	Q9ZKS9 mycoplasma
39	171	4.5	1176	2	Q9F231 helicobacte
40	170.5	4.5	1197	2	Q45888 clostridium
41	170	4.5	1837	3	Q74424 schizosacch
42	170	4.5	2752	5	Q9BJX0 plasmodium
43	169.5	4.5	786	5	Q77357 plasmodium
44	169.5	4.5	1344	2	Q49545 mycoplasma
45	169.5	4.5	2340	16	Q9ZD91 rickettsia

ALIGNMENTS

RESULT	ID	Q937W2	PRELIMINARY:	PRT:	741 AA.
AC	Q937W2				
BT	01-DEC-2001	(FREMURREL, 19, Created)			
BT	01-DEC-2001	(FREMURREL, 19, last sequence update)			
BT	01-DEC-2001	(FREMURREL, 19, last annotation update)			
DE	PAG PROTEIN (FRAGMENT).				
GN	PAG.				
OS	Bacillus anthracis.				
OC	Plasmid pX01.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;				
OC	Bacillus/Staphylococcus group; Bacillus.				
OX	NCBI_TaxID-1392;				
OX	[1]				
RP	SEQUENCE FROM N.A.				
RC	SPRAIN-FERRAR:				
RA	Adone R., Pasquali P., La Rosa G., Marianelli C., Muscillo M.,				
RA	Fasanello A., Francia M., Ciuchini F.;				
RT	*Sequence analysis of the genes encoding for the major virulence				
RT	factors of bacillus anthracis vaccine strain 'Carbopap' *;				
RU	Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.				
DR	EMBL: AJ413937; CAC93935.1; -				
KW	Plasmid.				
FT	NON TER				
FT	NON TER				
SQ	SEQUENCE	741 AA; 83153 MW; C7P95820E73065C0 CRC64;			

Query Match	98.0%; Score 3702; DB 2; Length 741;
Best Local Similarity	99.7%; Pred. No. 2.2e-176;
Matches	720; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY	1 EVKQENRLNLSSESSQGLGTFPSDNLFOAPVYVTSSTGDLSPSSLELNFSENGYR 60
DB	20 EVKQENRLNLSSESSQGLGTFPSDNLFOAPVYVTSSTGDLSPSSLELNFSENGYR 79
QY	61 QSAIWSGRTKKKSDYETPATSADNHYTMVDDQEVINKASNSKIRLEKGLVQIKIY 120

```

Db      80 QSAIMSGFIKKKSDSEYTFATSDADNHVMMVDDQEVYINKASNSKIRLEKGRLYOIKIY 139
QY      121 QRENPEKGLDFKLYMTDSQNKKEVYISDNLOLPELKOKSSNRKRSAGPTVPDDN 180
Db      140 QRENPEKGLDFKLYMTDSQNKKEVYISDNLOLPELKOKSSNRKRSAGPTVPDDN 199
QY      181 DGIPLSEYEGYTVDYVKNRFTLSPWISNIHEKGLTYKSSPEKNSADSDYSDPEKYT 240
Db      200 DGIPLSEYEGYTVDYVKNRFTLSPWISNIHEKGLTYKSSPEKNSADSDYSDPEKYT 259
QY      241 GRIDKNSPEARHPDLVAAPPIVHVDMENIILSKNEOSTONTDSEFTTSKTSRTH 300
Db      260 GRIDKNSPEARHPDLVAAPPIVHVDMENIILSKNEOSTONTDSEFTTSKTSRTH 319
QY      301 SEVHGNAEYHASFEDIGGSVAGSFNSNSTVAIDHSLSLAGEPTNAETMGLNTADTARL 360
Db      320 SEVHGNAEYHASFEDIGGSVAGSFNSNSTVAIDHSLSLAGEPTNAETMGLNTADTARL 379
QY      361 NANIRVYNTGAPLYNVLPFTSLVYGNKOTLATIKAKENOLSQLAPNNYPSKMLADIA 420
Db      380 NANIRVYNTGAPLYNVLPFTSLVYGNKOTLATIKAKENOLSQLAPNNYPSKMLADIA 439
QY      421 LNAODFSSPTITMNTNNOFLEKTKOLRLDTPQVGNATYENENGRVAVDTGSMSSEV 480
Db      440 LNAODFSSPTITMNTNNOFLEKTKOLRLDTPQVGNATYENENGRVAVDTGSMSSEV 499
QY      481 LPOIOETTRARIIFNGKDLNVERRIAANPSDPLETTKPDMTLKEALKIAFGFNPNCNL 540
Db      500 LPOIOETTRARIIFNGKDLNVERRIAANPSDPLETTKPDMTLKEALKIAFGFNPNCNL 559
QY      541 OYQKDIETEFNPDQOTSQNIKNOLAELNATNIYVLDKIRLAKANNILIRDKRFHNR 600
Db      560 OYQKDIETEFNPDQOTSQNIKNOLAELNATNIYVLDKIRLAKANNILIRDKRFHNR 619
QY      601 NNIAVGADESYYKAEHREYVINSSTEGLLNIDKIRKILSGIYEIEDTGLKEVINDRY 660
Db      620 NNIAVGADESYYKAEHREYVINSSTEGLLNIDKIRKILSGIYEIEDTGLKEVINDRY 679
QY      661 DMLNISTLRQDKTFIDFKKYNKDLPLXISNPNYKVNVAATKENTLINSENGDTSTNG 720
Db      680 DMLNISTLRQDKTFIDFKKYNKDLPLXISNPNYKVNVAATKENTLINSENGDTSTNG 739
QY      721 IK 722
Db      740 IK 741

```

```

SQ      SEQUENCE 743 AA: 83363 MW: 292757AD2D5D55A6 CMC64:
Query Match      98.0%; Score 3702; Db 2; length 743;
Best Local Similarity 99.7%; Pred. No. 2,2e-178;
Matches 720; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY      1 EVKQENRLNLSSESSSSGGLGYFSDLNFOAPMVYTSSTGDLSPSELEINPSENGYF 60
Db      22 EVKQENRLNLSSESSSSGGLGYFSDLNFOAPMVYTSSTGDLSPSELEINPSENGYF 81
QY      61 QSAIMSGFIKKKSDSEYTFATSDADNHVMMVDDQEVYINKASNSKIRLEKGRLYOIKIY 120
Db      82 QSAIMSGFIKKKSDSEYTFATSDADNHVMMVDDQEVYINKASNSKIRLEKGRLYOIKIY 141
QY      121 QRENPEKGLDFKLYMTDSQNKKEVYISDNLOLPELKOKSSNRKRSAGPTVPDDN 180
Db      142 QRENPEKGLDFKLYMTDSQNKKEVYISDNLOLPELKOKSSNRKRSAGPTVPDDN 201
QY      181 DGIPLSEYEGYTVDYVKNRFTLSPWISNIHEKGLTYKSSPEKNSADSDYSDPEKYT 240
Db      202 DGIPLSEYEGYTVDYVKNRFTLSPWISNIHEKGLTYKSSPEKNSADSDYSDPEKYT 261
QY      241 GRIDKNSPEARHPDLVAAPPIVHVDMENIILSKNEOSTONTDSEFTTSKTSRTH 300
Db      262 GRIDKNSPEARHPDLVAAPPIVHVDMENIILSKNEOSTONTDSEFTTSKTSRTH 321
QY      301 SEVHGNAEYHASFEDIGGSVAGSFNSNSTVAIDHSLSLAGEPTNAETMGLNTADTARL 360
Db      322 SEVHGNAEYHASFEDIGGSVAGSFNSNSTVAIDHSLSLAGEPTNAETMGLNTADTARL 381
QY      361 NANIRVYNTGAPLYNVLPFTSLVYGNKOTLATIKAKENOLSQLAPNNYPSKMLADIA 420
Db      382 NANIRVYNTGAPLYNVLPFTSLVYGNKOTLATIKAKENOLSQLAPNNYPSKMLADIA 441
QY      421 LNAODFSSPTITMNTNNOFLEKTKOLRLDTPQVGNATYENENGRVAVDTGSMSSEV 480
Db      442 LNAODFSSPTITMNTNNOFLEKTKOLRLDTPQVGNATYENENGRVAVDTGSMSSEV 501
QY      481 LPOIOETTRARIIFNGKDLNVERRIAANPSDPLETTKPDMTLKEALKIAFGFNPNCNL 540
Db      502 LPOIOETTRARIIFNGKDLNVERRIAANPSDPLETTKPDMTLKEALKIAFGFNPNCNL 561
QY      541 OYQKDIETEFNPDQOTSQNIKNOLAELNATNIYVLDKIRLAKANNILIRDKRFHNR 600
Db      562 OYQKDIETEFNPDQOTSQNIKNOLAELNATNIYVLDKIRLAKANNILIRDKRFHNR 621
QY      601 NNIAVGADESYYKAEHREYVINSSTEGLLNIDKIRKILSGIYEIEDTGLKEVINDRY 660
Db      622 NNIAVGADESYYKAEHREYVINSSTEGLLNIDKIRKILSGIYEIEDTGLKEVINDRY 681
QY      661 DMLNISTLRQDKTFIDFKKYNKDLPLXISNPNYKVNVAATKENTLINSENGDTSTNG 720
Db      682 DMLNISTLRQDKTFIDFKKYNKDLPLXISNPNYKVNVAATKENTLINSENGDTSTNG 741
QY      721 IK 722
Db      742 IK 743

```

```

RESULT 3
ID      032739 PRELIMINARY: PRT; 876 AA.
AC      032739;
DT      01-JAN-1998 (TREMBlrel. 05, Created)
DR      01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DE      ADP-RIBOSYLTRANSFERASE.
GN      CDTB.
OS      Clostridium difficile.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
CX      NCBI_Taxid=1496;

```

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C0196;
 RA MEDLINE-97230316; PubMed-9119480;
 RT "Perille S., Gilbert M., Bourlioux P., Corthier G., Popoff M.R.;
 RL "Production of a complete binary toxin (actin-specific Adp-
 DR HSP; P13423; IACC.
 DR HSP; P13423; IACC.
 DR InterPro: IPR003896; BinaryTox.
 DR PRINTS; PRO1391; BINARYTOXIN.
 KM Transferase.
 SQ SEQUENCE 876 AA: 98797 MW: 256062D45CE2B3B CRC64;

Query Match 23.6%; Score 890.5; DB 2; Length 876;
 Best Local Similarity 31.8%; Pred. No. 5,7e-37;
 Matches 253; Conservative 139; Mismatches 278; Indels 125; Gaps 30;

4 QENLLNESSSSGGLGYFSDLNQAPMVTSTTGDLSPSELENIPSENOYFOS 62
 40 KKKIYMEDILFNNGLMGTIFEDHKKDLKMAPIKDKNGLEKKEKVDLKKDKSDVAS 99
 63 AMSGFIKKKDETFATSDNHTVMVDOEVINKASNSKIFLEKGRLYQIKIYOR 122
 100 IRMGRIIPSKGDEYTLSTROD-VLMQVNESTI--SNTLKVMKKGKRYKRIEIO 155
 123 EN--PTKGLDFKLYWDSQNKKEVISDNQLPELKOKSSNRKSTAGPTV--- 176
 156 KNGISDNLSPLWY-ELDGMKKIIPENLELDY---SNEKD---DPFIPNNNF 205
 177 -----DRNDGIPDSLEVBGYVDVKNKRFELSPWISNHEKKGLTYKSS 222
 206 FDPRLMSDEWEDLDTPDNDNIPDSYERNQYTI---DLIAVAKMDSFAE-OGKKKYSVN 260
 223 PEKMSYASDPYDEFEKATGRIDKNVSPKARHVAAPYVHVDMENILSKNEDOST 282
 261 YLESTAGDPYDYDKAGSGFKAITEARPDVAAPYVGVGMKLLISTNEHAST--- 317
 283 DSERTISKNTSTSTSTSEVHGNAEVAHASFDDIGSVSAGFSNSSTVAIDHS----- 337
 318 -DQKTYGSRATNKSKESTNA-----GSAVNVQVQNGFTANVTNHTSTTON 363
 338 ---LSLAGEPTAETMGANTADARLANIRVYNGTAPYVNLPTTSLVGGKNOTLARI 394
 364 STAVQDSNGESMTGISTNGESYINANVRYNGTAPYKATPTMLV-DGDLTLSTI 422
 395 KAKENQISQILAPNNYTPSKNLAPILNAADDESSPTITMNTNOLELEKTKQLRLTDQ 454
 423 KAKENQISQILAPNNYTPSKNLAPILNAADDESSPTITMNTNOLELEKTKQLRLTDQ 482
 455 VYGNIAIYFNGRVRVDTGSNNSEVLPOIOTETARIIFNGKDLNVERIAAVNSDPL 514
 483 VSGNFGTKN-SSQI-VTEGNSMSDVISQIDISASITLIDTEN-ESYERRVTAKNLQDE 539
 515 ETTPEDTTLKALKAIFGNEPENGNOYOGKQITE--FDFNEQOOTSONIKNOIAELNAT 572
 540 DKT-PELTIGELIEKAFGATKKGGLYFNDIPIDESCVELIFEDPTANKIKDSIKLISDK 598
 573 NIIVYLDKIKLAKANNILIRDKRFHY---DRNNIAGVADSVYKFAHVEVNSSTGL-- 627
 599 KIYVNV---KLEGNMILIKPTPTFFNFDYNNP--STMSVNTTNOGGLOGSANKLNG 652
 628 -----LMIIDOKRLKILSGY-----IYEI---EDDEGI-----KEYI 656
 653 EFKIKIPMSLAKRYKRYVSGYSKDLVTSNLSYIKAKKEKMDLVLEOGYTKFYERE 712
 657 NDRDMNTS-SLRDQKTFIDFKRYND--KLPLYSNPKYK-----NVYAVYKENT 706
 713 TTEKDSNIEITLIGSGTYIDNLSITELNSPELIDEPYKIPFDQIDMAKHTIPADL 772
 707 IINSENGDTSTNGI 721

DB 773 NFNP-TCNTYINCA 786

RESULT 4
 ID 006498 PRELIMINARY; PRT; 879 AA.
 AC 006498;
 DT 01-JUL-1997 (TREMUREL. 04, Created)
 DT 01-JAN-1998 (TREMUREL. 05, Last sequence update)
 DT 01-DEC-2001 (TREMUREL. 19, Last annotation update)
 DE SB COMPONENT.
 GN SBS.
 OS Clostridium sporforme.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes.
 OX NCBI_TaxID=29348;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C0246;
 RA Gilbert M., Perille S., Daube G., Popoff M.R.;
 RT "Clostridium sporforme toxin genes are related to C. perfringens iota
 RL toxin genes but have a different genomic localization";
 RT Syst. Appl. Microbiol. 20:357-347(1997).
 DR EMBL; X97969; CAAB612.1;
 DR HSP; P13423; IACC.
 DR InterPro: IPR003896; BinaryTox.
 DR PRINTS; PRO1391; BINARYTOXIN.
 SQ SEQUENCE 879 AA: 98739 MW: 40685ACB8E05BA01 CRC64;

Query Match 23.6%; Score 890.5; DB 2; Length 879;
 Best Local Similarity 28.9%; Pred. No. 5,7e-37;
 Matches 261; Conservative 147; Mismatches 257; Indels 237; Gaps 34;

4 QENLLNESSSSGGLGYFSDLNQAPMVTSTTGDLSPSELENIPSENOYFOS 62
 36 QEVNTNEKTVASGNGYGFADHKKDLKMAVKKNGLEKKEKVDLKKDKSDVAS 95
 63 AMSGFIKKKDETFATSDNHTVMVDOEVINKASNSKIFLEKGRLYQIKIYOR 122
 96 IRMGRIIPSKGDEYTLSTROD-VLMQVNE--GEIANTLKVMIKQGYSTRIEIO 150
 123 ENPEKGLDF-----KIYWDSONKKEVISDNQLPELKOKSSN----- 162
 151 ----DNDIGYVDLSSPKLY-ELNDGKTLIPKKNLEFLNDYSKIDENPFIPKDNFEDK 205
 163 --SKKRSTSA-GPTVPDRNDGIPDSLEVBGYVDVKNKRFELSPWISNHEKKGLTYK 219
 206 LKSRASARLASGDEDEDLTPDNDNIPDAIEKNGITL-----KDSIAVKN-EDSPKQGYKRY 260
 220 KSPKMSYASDPYDEFEKATGRIDKNVSPKARHVAAPYVHVDMENILSKNEDOST 279
 261 LSTYLESNAGDPYDYDKAGSGFKAITEARPDVAAPYVGVGMKLLISTNEHAST 320
 280 QNTDSERTISKNTSTST--HTSEVHGNAEVAHASFDDIGSVSAGFSNSSTVAIDHS 337
 321 ----DQKTYGSRATNKSKESTNA-----GSAVNVQVQNGFTANVTNHTSTTON 363
 338 LSLAGEPTAETMGANTADARLANIRVYNGTAPYVNLPTTSLVGGKNOTLARI 397
 374 ---GCE-SWNTSLSINKGESYINANVRYNGTAPYKATPTMLV-DGDLTLSTI 428
 395 KAKENQISQILAPNNYTPSKNLAPILNAADDESSPTITMNTNOLELEKTKQLRLTDQ 457
 423 KAKENQISQILAPNNYTPSKNLAPILNAADDESSPTITMNTNOLELEKTKQLRLTDQ 482
 429 DNQGNLSNENYTPSKNLAPILNAADDESSPTITMNTNOLELEKTKQLRLTDQ 488
 458 NIIVYLDKIKLAKANNILIRDKRFHY---DRNNIAGVADSVYKFAHVEVNSSTGL-- 515
 489 N---YGIKSGOQIITBEGNSMSDYISQIDISASITLIDGSD--VEERVYAVDSSNPED 543
 516 TTKEDSNIEITLIGSGTYIDNLSITELNSPELIDEPYKIPFDQIDMAKHTIPADL 573
 544 KT-PVLTIGELIEKAFGATKKGGLYFNDIPIDESCVELIFEDPTANKIKDSIKLISDK 602

```

OY 574 IYTVLDKIKLAKNANILIRD----- 593
DB 603 IYVY-----QLEGRKILIKITSTYFNNFNGYNNPSSKSNVNDNODLOANAKLSGETK 658
OY 594 -----QLEGRKILIKITSTYFNNFNGYNNPSSKSNVNDNODLOANAKLSGETK 598
DB 659 IYVPMKLNPKRKYVSGYLKSNSTNPTIVNIKAEQTYVLYSENDYKRFSEFEETIG 718
OY 599 -DRNN-----IYVPMKLNPKRKYVSGYLKSNSTNPTIVNIKAEQTYVLYSENDYKRFSEFEETIG 620
DB 719 RDSNNEITLSSGTFIFLNDLSTELNSTPEILKEPDIKVPDDELL-DAHKKYADLSF 777
OY 621 NSSTEGILLN-----IDKDIRILSGYIVLEL-DTEGLEVYINDRYMLNLSLROD 671
DB 778 NOSTANYIYDGLYEPYQTKREVLIDYIOKRYEATLEYSGFDICTOKDELRYNGTDSNQ 837
OY 672 GKT-FIDEKYNKPLIYISNPNY-----KVNVAVTRENTIINPENGDTSTGIRKILI 726
DB 838 PKTNYVNFYSY-----FTSGENVMPYKKRIYALIPEN-----KELLY 875
OY 727 FS 728
DB 876 LS 877

```

```

RESULT 5
OY 09KHA1 PRELIMINARY; PRT: 876 AA.
AC 09KHA1;
DB 01-OCT-2000 (TReMBLrel. 15, Created)
DB 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DB 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DB CDB.
GN CDB.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-CCUG 20309;
RA Chang S.Y., Song K.P.;
RT "ADP-ribosylating Binary Toxin Genes of Clostridium difficile strain
RT CCUG 20309."
RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: A8271719; A881761.1;
DR HSP: P13423; JACC.
DR InterPro: IPR003896; BinaryTox.
DR PRINTS: PR01391; BINARYTOX1B.
SQ SEQUENCE 876 AA; 98793 MW; 366D629352E745A5 CRC64;

```

```

Query Match 23.5%; Score 888.5; DB 2; Length 876;
Best Local Similarity 31.7%; Pred. No. 7.2e-37;
Matches 232; Conservative 140; Mismatches 278; Indels 125; Gaps 30;
OY 4 QENRLNSESSESSGGLLYGYSDFLNFQAPVYVTSSTFGDLSIPSSSELEN-IPSENOYFOS 62
DB 40 KKEIYENEDILPPNNGMLGYFTDEHFKDLKMAIPKIDGMLKKEEKYKDLDDKSDVKS 99
OY 63 AMSGIKVKSDEYFATGADNHTYVAVDDEVIKAKSNKIKLEKGLVQIKIORY 122
DB 100 IRTGRIIISKDGTYLTSTDRD-VLMQVNTESI---SNILKVMKMGKREYVRIEOD 155
OY 123 EN--PEKGLDFKLYMTDSONKKEVYISDNLOJPELKOKSSNRKSTKSTAGTYV--- 176
DB 156 KNIGSIDNLSNPNLYW-ELDKMKKILIPENLFLRDY---SNJEKD-----DPEIPNNF 205
OY 177 -----DRNDGIPPSLEVGTYVYVKKRFTPLSPMSIHKKGLTKRYS 222
DB 206 FDKLMSDMEDEDDLDNDNDIPDSYERNGYTI---KDLINKMEDEFAE-GQYKRYYSN 260

```

```

OY 223 PEKSTADPYSDDEKTYGRIKKNVSPKARPLVYAVPIVYVDMENILISKNEOSTONT 282
DB 261 YLESNADPPTDYDEKASGDFKAKTEADPLVAVYVGVGKELISTEHAET--- 317
OY 283 DSEFTIKSTSTSRHTSEVHNAEVSASFEDIGSASGFSNSSTVADHS----- 337
DB 318 -DQKTVSRATNTKTESNTA-----GVSNVGYONGFTANTYNSHTDN 363
OY 338 ---LSAERTMAETMGLTADTARLANIRYVNTGTAPITNV-LPTSVLKRNQTLATI 394
DB 364 STAVDPSNGESMNGSLINKGESAVIANVRYNGTAPMYKVPPTNVL-DGPTLSTI 422
OY 395 KAKENOLSOILAPNNYPSKNIAPALNAODFSSTPIYNNYNOJLEKTRQRLDQ 454
DB 423 KAENQIGNNLSGPDYTPKKGSLPALNTMDGFSRLPIVNDLKKLADGROIKLETQ 482
OY 455 VYGNIAIYVNFNGSVRYVDTGSNMSEVLPJOIETARIFPNKDLNVERIAVNSDPL 514
DB 483 VSGNFQTKN-SSGQI-VTEGNSMSDYISQDISASITLDTEN-FSEYERTAKMLDPE 539
OY 515 ETKPMTEKALKIAGFENPNSMLOQKDTL-EDFNDQJNSIKKQALAEAT 572
DB 540 DKT-PEITIGAIKAPGATKDGILFNDIPIDESCCELLFDRHANKIKDLSLSDK 598
OY 573 NIYTVLDKIKLAKNANILIRDKRPHY---DRNNIYVAGDESUYKAREVINSSTEG-- 627
DB 599 KIYVY-----KLERGMNLIKRTYPTNPDYNNY-STSNSVNFNKDGLGSKANKING 652
OY 628 -----LNLIDKIRILSGY-----IVEL---EDTELT-----KEVI 656
DB 653 ETKIKIPSEIKPKRYVSGYKNDPLNSIIVKIKAKKEKTYLVAPDGYTKPSYEF 712
OY 657 NDYVDMNLS-SLRDQKTFIDPKYND--KLPYISNPNYK-----VNYVAETENT 706
DB 713 TTEKDSNIEITLGSSTYLDNLSTELNSTPEILDEPEVKIPJOEILDAHKIYPADL 772
OY 707 IINPENGDTSTNGI 721
DB 773 NFNSP-TGNTYINCM 786

```

```

RESULT 6
OY 046221 PRELIMINARY; PRT: 875 AA.
AC 046221;
DB 01-NOV-1996 (TReMBLrel. 01, Created)
DB 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DB 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DB IOTA TOXIN COMPONENT IB PRECURSOR.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-NCIB 10748;
RA MEDLINE-94041637; PubMed-8225592;
RA Perelle S., Gilbert M., Bogue P., Popoff M.R.;
RT "Characterization of Clostridium perfringens Iota toxin genes and
RT expression in Escherichia coli."
RL Infect. Immun. 61:5147-5156(1993).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN-NCIB 10748;
RA Popoff M.R.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: X73562; CAA51960.1;
DR HSP: P13423; JACC.
DR InterPro: IPR003896; BinaryTox.
DR PRINTS: PR01391; BINARYTOX1B.
KV SIGNAL.
FT SIGNAL. 34 38 POTENTIAL.
FT CHAIN 212 875 IOTA TOXIN COMPONENT IB.

```


RESULT 8
ID 093TV6 PRELIMINARY; PRT: 2276 AA.
AC 093TV6;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE BIOPFILM-ASSOCIATED SURFACE PROTEIN.
GN BAP.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-V329;
RX MEDLINE=21189316; PubMed=11292810;
RA Cucarella C., Solano C., Valle J., Amorena B., Lasa I.,
RA Penades J.R.;
RT "Bap", a Staphylococcus aureus Surface Protein Involved in Biofilm
RT Formation.
RL J. Bacteriol. 183:2888-2896(2001).
DR EMBL: AF288402; AAK38834.1;
SQ SEQUENCE 2276 AA; 238792 MW; 8623C60472CB0C0F CRC64;

Query Match 5.3%; Score 200.5; DB 2; Length 2276;
Best Local Similarity 19.2%; Pred. No. 0.094; Mismatches 300; Indels 267; Gaps 39;
Matches 166; Conservative 131;

47 SSELINPSENGYFQSAIWSGFIYKKSDEYTFATSDNHVYMWVDOQEVINKASNSKI 106
DB 43 AAEIDITTEDVQDGDGEALDQIKNKSEKDYTEDDNNAYVQNSAQYVDSSENSDT 101
QY 107 RLEGRGLYQIKYQYQREMPREKGLDFKLYTPDSQNK-----VISDMLQPELK 157
DB 102 AVESTN-DSVKDEKTESKNSAO-----DDNKESSNQEESTVTSOSSSEVPQK 154
QY 158 QKSSSKKRSKRSAGPYPP---RONDGIPDSLEYGVYVYKRRFLSPWISNHEKK 214
DB 155 -KQNETSETALDEASTKQNNKNDNRADDNDIKEDS-NQOESTVTSOSSSEVPQK 212
QY 215 GLTYKSSPEKMSA-----SDPYSPFEKVTGRIDKVSPEARH 253
DB 213 -----KQPKSSSKIKPKQOEEVAKKEKAITETLADNKKELTKNNKTDNKESELE- 266
QY 264 PLVAAPYIVHVMENIILSKNEDQ-----STQNTDSETRISKNTSTSRTHSEVHGA 307
DB 267 -----SNLSSENKKDVTESFLNSQLSDSEKTKIMEANADYDKATDEEINT 313
QY 308 EV-HASFPIGGSVSAGFSNSNSTVAIDHSLSLAGRTMAETMGLNPTADTARLANIRY 366
DB 314 ELRLASLIEH-----ANNKKRTETL-----ATQQRTPFAATPTALRAAVNODEEL 360
QY 367 VMTGTAIYVNPPTSLVLGKNOFLATIKAKENL-----SQT-----AP 407
DB 361 -----QKSLGYTDNTPTASMLPDPKGLSDDALNSNIIPDHLISYMSGANG 407
QY 408 NNYPSKNIAP-----ALNAODPFSSPTI--TMVNOFLLEKTKOLRL--DTPQVYG 457
DB 408 NRYKIDKLDPILIAHYTKISANPSGSKPVEFARKKNDGNLDTQWVNFPIANDGLFG 467
QY 468 N--IATVNEKGRVAVDTGSNMSEVLQIOETARTIIFNGDL-----NLVE 502
DB 468 GAEILSGYAKNGKIELD-----DIVGNISNAONLSNNKLNHQVFVADSE 514
QY 503 RRIAAVSPDPLETKRPMPL-----KALKIANG--FNEPKNL----- 540
DB 503 RRIAAVSPDPLETKRPMPL-----KALKIANG--FNEPKNL----- 540
QY 515 NKIVRSSESGYFLKADDDVLLENVSTENNNAFASSGSGATYNNVEVPEGILIIQQ 574
DB 515 NKIVRSSESGYFLKADDDVLLENVSTENNNAFASSGSGATYNNVEVPEGILIIQQ 574
QY 541 -----QYGGDIIEEDFNFDDQ----- 557

DB 575 IMKNGIFSRTKANQWAVYQIDKDLPIYEGVELHGYDYKGLNGCDKNYAKKRVADL 634
QY 558 -----TSQNKIKQMLAEANNTIYVLDKI--KINAKM--NIIIRKRPBYDRNIA 604
DB 635 TIDEVNGNTITSDL-NKLIERNNMLPETYGAVYLAIKNSVNNILTKDAKDSGN--- 690
QY 605 VGADSEYVKEKAREVIN-----STREGILLIMDKIRILSYLVEIEDPEGIAKVIN 657
DB 691 -----LIRETKKEDPTFPAGYLDKSGALIN-----NTLGSYLAQDYD--KQGL 736
QY 658 DRID-MLNISSLRODGKTFIDPKKYNDKLPYISNPNT-KVNVAVYKENTIIINS---- 711
DB 737 DREKQLSLDAENE-DTQGDKNQDGEVNVYNTSPVGKPOAADITTEDVAVSGSVPLK 795
QY 712 ENGDTSTNGIKRILIFSRRGYEIG 735
DB 796 EGAAQT-----AKVINAGETVVG 814

RESULT 9
ID 09P008 PRELIMINARY; PRT: 4688 AA.
AC 09P008;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE HYPOTHETICAL PROTEIN U0482.
GN U0482.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmatetaceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROVAR 3.
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA Caselli G.H.;
RT The complete sequence of the mucosal pathogen Ureaplasma
RT urealyticum.
RL Nature 407:757-762(2000).
DR EMBL: AF002145; AAF30894.1;
DR InterPro: IPR001152; Thymosin_B4.
DR InterPro: IPR000626; Ubiquitin.
DR Pfam: PF00240; Ubiquitin_1.
DR SMART: SM00152; THY; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 4688 AA; 534880 MW; B53ABFAFE1997F CRC64;

Query Match 5.3%; Score 199; DB 16; Length 4688;
Best Local Similarity 21.2%; Pred. No. 0.29; Mismatches 325; Indels 186; Gaps 40;
Matches 173; Conservative 132;

QY 19 ILGYPSDNLQAPMVVTSSTGDISPSELENISENOYFQSAIWSGFIYKKSDEY 78
DB 19 ILGYPSDNLQAPMVVTSSTGDISPSELENISENOYFQSAIWSGFIYKKSDEY 78
QY 3699 LVQVYLD-NIHONIDEPTRIKPDHNV-SKEIINPGVYMIKSGNNKSPDTTANPEFK 3756
DB 3699 LVQVYLD-NIHONIDEPTRIKPDHNV-SKEIINPGVYMIKSGNNKSPDTTANPEFK 3756
QY 79 FATSADNHYTWVDOEVINKASNSNKRILEKGRLOYKIOYORENPTKGLDFKLYWD 138
DB 79 FATSADNHYTWVDOEVINKASNSNKRILEKGRLOYKIOYORENPTKGLDFKLYWD 138
QY 3757 IETQ-----DDNDVNLNIDATVYFKDEHNINIKOKIIRIKEN-----ND 3795
DB 3757 IETQ-----DDNDVNLNIDATVYFKDEHNINIKOKIIRIKEN-----ND 3795
QY 139 SQNKKREYISDNQLPELKQSSN-----SRKKRSTSGAPTPVPRD----- 179
DB 139 SQNKKREYISDNQLPELKQSSN-----SRKKRSTSGAPTPVPRD----- 179
QY 3796 WLKQGI-----DNLM-PETVYKLENIELSKPLKTHNLSVSIENDKENISLITETGPNVLYX 3851
DB 3796 WLKQGI-----DNLM-PETVYKLENIELSKPLKTHNLSVSIENDKENISLITETGPNVLYX 3851
QY 180 -----NDGIPDSLEYGVYVYKRRFLSPWISNHEKKGLTYKSSPEKMSASDPYSD 225
DB 180 -----NDGIPDSLEYGVYVYKRRFLSPWISNHEKKGLTYKSSPEKMSASDPYSD 225
QY 3852 IGQONDITNDPQOTINVTLSGVNSK-YNGRQIKVYVYKONNVYIYES--LITLQKKN 3907
DB 3852 IGQONDITNDPQOTINVTLSGVNSK-YNGRQIKVYVYKONNVYIYES--LITLQKKN 3907
QY 236 FEKVTGRIDKVSPEARHPLVAAPYIVHVMENIILSKN-EDOSTQNTDSEF-RTISKRN 293
DB 236 FEKVTGRIDKVSPEARHPLVAAPYIVHVMENIILSKN-EDOSTQNTDSEF-RTISKRN 293

```

Db 3908 YOLLNLNSN-----REYFEKIEINHSNTNNFEDLEKNGVSNFTIQTQNT 3957
QY 294 STERTTS-EVHGNAEYHASF-----PDIGGSAGFS-----NSMS---STVAD 335
Db 3958 TVOMNDSATTVIGRGVNEFKIKSEKILLENNOQVAFKAKETIDYVIMLQYTRPLK 4017
QY 336 HSLSLAGETAAETMGLNT---ADTAELMANIKRYVTGATPIYVLPFTTSLVAGKNOTLA 392
Db 4018 DVTSDKEBTMAHDLNSVNFKEETTYLVKIOFVFKPKAKANNINSENNVILDTNSI 4077
QY 393 -----TIKAEENOLSOILAPNNYPSKNLAPALN-ADODESSTPTMYN-NOFLELE 443
Db 4078 NSVFEETTVGDHKLINTTSSNNVNTNQTINFTLSGVKMSVGRKTKLSYSNPTJSSI 4137
QY 444 KTRQLRLTDQYGNITATYVNFENGR-----VYDGSNMSEVLPJOIETTARIIFPNKD 497
Db 4138 HTEVVLIESNKTQVNIILNKLKRNFTYLLIDVKLIDNNVSDFKECENLTNSPTTSTA 4197
QY 498 LNLVERRIAANVPSDPLETT-----KPDWTLKEA-----LKIAFGNEPNCMLQ 541
Db 4198 INVNLTEISNASTNLKSTIKILNLPDNLVLDKQATYVGNKQAMGFIVSGNIK 4257
QY 542 YOGKDITEFDNFDOOTSONIK-NOLAEATNITYVLDKIKLAKANNILIRDKRFHYD- 599
Db 4258 YLATVLDLNFN-DKVNIVISFNKRPISIAEN-----IGDKSNITL-----YNDNS 4304
QY 600 -----RNINIVGA-----DSVYKFAHREVINSSTEGILLNIDKQIRKILSGYVIEDETE 650
Db 4305 IPRLEINNDIIVNGPIKKEIYVKNANK-NNIDVDLGQINPRILAHNR-PLAFKSTN 4361
QY 651 GKEVINY-RYDMLNITSL-RODGKPEIDF-----KTY-----681
Db 4362 -----NOIEIVNGSSIVNNDKSTIRFTPLANKAKLXSLVDYIYVNNNSNTIVE 4415
QY 682 NDLPLXISNRYVNVYAVTAKENTIIINPSENDTS 717
Db 4416 SNKLP-KLNNINITYOK--INSHITL--SKNGEMS 4445

RESULT 10
QY04X0 PRELIMINARY; PRT; 2647 AA.
AC Q904X0;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PUTATIVE ERYTHROCYTE BINDING PROTEIN RBL-1 (FRAGMENT).
GN RBL-1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OC NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20078864; PubMed=10613703;
RA Peterson D.S.; Wellem T.E.;
RT "RBL-1, a putative erythrocyte binding protein of Plasmodium falciparum, maps within a favored linkage group in two genetic crosses."
RL MOL. Biochem. Parasitol. 105:105-113(2000).
DR EMBL; AF131999; AAB3018.1; -.
DR InterPro; IPR002828; Sure.
FT NON_TER 2647
SQ SEQUENCE 2647 AA; 304550 MW; AE98F8BD754E300 CRC64;

Query Match 5.0%; Score 188; DB 5; Length 2647;
Best Local Similarity 21.4%; Pred. No. 0.49;
Matches 183; Conservative 134; Mismatches 295; Indels 242; Gaps 46;
QY 1 EVKQERRLN-ESSSSQGLGYPSDLN-----FOAPM-----VYTSSTGD 42
Db 1080 ELREESPLDKTKTSAIG-----DKNHSYKSAIDIQSHINSNDRIAYSSVQVD 1132

```

```

QY 43 LSPSELEINPSPNOPOSATMGFLKVKSDYEYFATSDAHVIMWDDQEVY----- 97
Db 1133 SSSSSKSTESIRIDNDKDFNT---SEDIAPINHEKIGSSADKCS---EDKSTIIDSE 1186
QY 98 ----NKASNSKIRLE-----KGLYQIKIOYORNPETKGLDRLKLTWD 138
Db 1187 NEBNKSSHSDIKQSDNCGSTIDYSLSEESPQDLESY-----SPSSIDDLK----- 1234
QY 139 SONKREVIS---DNLQLEIKOKSSNRKRSSTAGPYPRDND-----GLPDS 186
Db 1235 -PKKSSPVTSFDVHDSFNISELQASQONADSYGKERPSKNLRTDDVSEKKSIVSPN 1293
QY 187 LEVEGYTVQKNNKTFELSPWISNIHERKGLTKYKSSPK-----WSTASDPDEFKY 239
Db 1294 VSY---TYDEBDKRGQISD-DSSIH-----HEIDPEKNLHV-SFSLSGLEBDIEKE 1342
QY 240 TGRIDKN---VSPERAPLVAAPIV-----HYDMENILSKNEDOSTONDSE 285
Db 1343 KGEKGSILPIPSPEKINDGKRENTVDPVSERVDNDRDIDSNV---SEDESSIIISR 1399
QY 286 TRTISKNTS---TSRHTSEVHGNAEYHASFIDGSVAGFS-----NSNSTVA 333
Db 1400 NTEGINSEFLKSEHTSVDNRRDNHKNQENLVSSSTQSESREKREKENADSS--- 1456
QY 334 IDHSLSLAGERTAEFMGLNTADTAR-----LNNIRYVNTGTA---PIYVNL-PTT 381
Db 1457 --HESELISISEVGETIRRNDAEASENDKEDILQSEGOVTEKTKLEPTVINLQST 1514
QY 382 SLVLRKNOTLATIKAKEN---OLSOILAPNNYPSKNLAPALNADODESSTPTMYNQ 438
Db 1515 PLGEIHKRSEIDDDNDGTGCVHTDANTFOYRBSSEVYITNKDPAKEMT--TKPESQ 1572
QY 439 FLELEKQALDLDQYVNT-ATYVFNENRVAVDTGSMSEVLVQIOE-----T 487
Db 1573 YIKKE-----IEIDSTKQDN---DQSN--SLIFLKNNEEDQVSTPS 1613
QY 488 TARIIPKNDLNLVERRIAANVPSDPLETTKPDNTKALKIAVENE-PNGLQYOGKD 546
Db 1614 TRVWMSG--SFVSRQEDIEEKDQKHIT-DDTINPSENGLKCGEMPNDSI---KS 1665
QY 547 ITEFDNFDOOTSONIKNOLAEATNITYVLDKI-KLAKANNIL-----IDRKR 595
Db 1666 VT-----ITESPRLDVEQMIPEIDPGKNEKNILIEPOESTTEIR-KQ 1707
QY 596 FHYDRNNTAVG-----ADESVYKFAHREVINSSTEGILLNIDKQIRKILSGYVE--- 645
Db 1708 MDGPIISNVNIPEELHPVAGSKLEEKERSMDADQK--TITEDITF-----VVEDPNG 1758
QY 646 IEDTEGLKEVINDRYDMLNLSLRDQKTFIDPKYINDKLPLYISNPNTKVVAVYTKEN 705
Db 1759 IGEHQNKLEVEHQASELNTYNSL-DGRTVVEYKERLIDNPGSLPNDRTTEHIELDREK 1816
QY 706 TIINPSE-----NGD 715
Db 1817 EIEHRELDANNGE 1830

RESULT 11
QY06216 PRELIMINARY; PRT; 2771 AA.
AC Q26216;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE RHOPTRY PROTEIN.
OS Plasmodium yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OC NCBI_TaxID=5861;
RN [1]
RP SEQUENCE OF 379-2771 FROM N.A.
RX STRAIN-YM;
RX MEDLINE=97077455; PubMed=8920022;
RA Sinha K.A.; Keen J.K.; Ogun S.A.; Holder A.A.;

```


FT NON_TER 2747 2747
SQ SEQUENCE 2747 AA; 321237 MW; 430508AC7B8624BD CRC64;

Query Match 4.8%; Score 183; DB 5; Length 2747;
Best Local Similarity 19.1%; Pred. No. 0.91;
Matches 162; Conservative 132; Mismatches 332; Indels 220; Gaps 36;

```

QY 1 EYKQENL-----LNESSSGGLGTYFSDLANFOAPVYVTSSTGDLSPSELENIPSE 56
DB 960 QTKLENFTEFSLNHNANNEELI-KTESDLAKALGINEENMLYNQFEKEKTPNDIKK 1018
QY 57 NOYQOSLNSGPIKAYKSDERTPATSDNHVYVWD-----QEVINKASNSNKTREK 110
DB 1019 NIHNEEISIKIEKTHAS-IYINISEETEREIGINIESLNTKVEEYKKNVYVTKR-EK 1076
QY 111 GRLYOI-----KIQY-----QRENTEKGLDFLWYVDSQNKKEVISS 148
DB 1077 LKHDFSDFGKGNIRKYDKIRKINDIMAVSQIDQIHIGLD-----DIORSSSEYVS 1130
QY 149 DNIQPELKQSSNSRKRKSTAGPTVPDRNDGIPDSLEVEGYTVYVKNKRTFLSPMS 208
DB 1131 EMEXKXKLEKVSNT-----EISND-----NVEG-----IKKQOII---VT 1164
QY 209 NIHEKKG-----LTQKSSPEKSTASDPYSDPEKYTG-----RIDKNVSPEAR 252
DB 1165 KIDKKKTYEEIKKLSEIKIEKDNLSLEKYDINISYQONIGMLFLEQIDEE-KKKA 1223
QY 253 HPLVAAPPIVHDMENILSKNEDOSTQVTDSETRTISKSTSTSRHTSEVHNAEVH-- 310
DB 1224 NTIKSMAYID-DLNDIKKKSQIEETENDIKMDINKEMALKISHDDKKCHDKSKNHE 1282
QY 311 --ASFPGIGSASGFSNSNSVAIDHSLSLAGERWALMGTLATATLANANRYV 368
DB 1283 NISDIYKSSKLIQOFNR-ESDINDIKMLQ-----KVVSSQNHNSDINOCL 1329
QY 369 TGTPIYNYLPTSLYLOKNOTLATIKKENOQLAPNNYPSK-----N 415
DB 1330 NEVANINIL-----KLNKIKIIDKYEYSEIKKKKNKINDELN 1371
QY 416 LAPIALAADOSSPTITMANTNOFL-----LEKTKOURLDTP-----QV 455
DB 1372 SEKVIKIKIBDLSLECKRSKINSTLDDCDECRKINIVLKNILNEETNTYXXFNAAE 1431
QY 456 YGNATYNEENGRVVDGTSNMSEVLPQIOETTARILFNGKDLNVERRIAANPSDPLE 515
DB 1432 YNKIVLSNENN---IEMADNKSQYITLTKRNNGT---NDHDYNTIKELK-SHKDKSNGTK 1483
QY 516 TTKPDMLKEALKIAFGNEPNCNLQYQGD-----ATEPDFNDQOTSQNIKNQ 565
DB 1484 -TEADOKKKAIOK-----NXLEFQYKEEYVTLNKKYAVELKNKFD-KTKNDSKOI 1533
QY 566 LAELNATNIYVLDKIKILNKKM-----ILIRPKRPHYDNNINIVAGDESVYKAEHREV 619
DB 1534 IREIKDANHCTLESCKSEKKMEIKNEKIHIEDEVANDKRSKMAITSIKVSEPKTKI 1593
QY 620 IN-----SFEGLLLMIDKIRKILSGYIETEDP-----BELKVINIDRYDMLNS 666
DB 1594 IKINEITSDDCKLEFNDELKOISMLSDQETKLENGKOLKLEEL-----LE 1644
QY 667 SLRODGTFIDKRYKNDKPLYSN-----PKYVAVYAKKENTJINSNGDTS 717
DB 1645 SLKOKKNIEDOKKLEDEVSKIKENTVBOHKKNIEIGI--VEKIMELAKTKKXKIES 1702
QY 718 TNGIKK 723
DB 1703 TKELIK 1708

```

```

DT 01-MAY-1999 (TREMUREL. 10, Created)
DT 01-MAY-1999 (TREMUREL. 10, Last sequence update)
DT 01-DEC-2001 (TREMUREL. 19, last annotation update)
DE LARGE SUPERNATANT PROTEIN 1.
CN LSPAL.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=35000;
RX MEDLINE=9030326; PubMed=9811662;
RA Ward C.K., Lumley S.R., Latimer J.L., Cope L.D., Hansen E.J.;
RT "Haemophilus ducreyi secretes a filamentous hemagglutinin-like
RT protein."
RL J. Bacteriol. 180:6013-6022(1998).
DR EMBL: AF057695; AAC79757.1;
DR InterPro: IPR000130; Zn_MTPepidase.
DR SMART: SM00235; Zmnc; 1.
SQ SEQUENCE 4152 AA; 456173 MW; 7082DEDC988AB8F3 CRC64;

Query Match 4.8%; Score 183; DB 2; Length 4152;
Best Local Similarity 22.2%; Pred. No. 1.6;
Matches 187; Conservative 118; Mismatches 321; Indels 218; Gaps 45;

QY 10 NESESSGGLGTYFSDLANFO-APVYVTSSTGDP-----LSIPSELENIPSENGYF 60
DB 162 SRRESSTQ-IVGKLANIOLQKEAKLIINQYTDHESNIOGALVPAOKADLIYV--- 216
QY 61 QSAIISGPIKAYKSDERTPATSDNHVYVWDQEVINKASN-----SKIKLEGRLY 114
DB 217 PNCITLNGKYITINDREYVTS-----DILPHREGLISVRNKVYIDRGVYA 264
QY 115 QIKIQY-----QRENTEKGLDFKIYVDSQNKKEVISSD-----NIQPELKQSS 161
DB 265 TNGLSFEVYARNIDQNG--KIYAKTENOKSYNPRANITTPAOSLNTNTEATPISS 321
QY 162 NSRKRKSTAGPTVPDRNDGIPDSLEVEGYTVD---VKNKRTFLSPWISNIHEKGLT 217
DB 322 GTSR--TSDDPAI-SADSAGSMTGSIKIFVYTDGAGVKKHGIIFENDINKMDGNA 377
QY 218 KYSSPEKSTASDPYSDPEKYTGRIIDKNVSPAPRPLVAAPPIVHDMENIILSKNEDQ 277
DB 378 SLKELYAK-----KDIDILANDIELETKQLOANNKIIILNSTGKIMLRNASEV 425
QY 278 STONTSETRTIS-KNTSTSRHTSEVHNAEVASFPGIGSASGFSNSNSVAIDH 336
DB 426 SADNVAVKSEINALENSMSANSLDVIYTKLEVNRS-----SKVSAGANIKKSNITIDG 480
QY 337 SLSLAGERTW-----AETMGLNTADTARLN-----ANIRYV---T 369
DB 481 SSVAVANKITLAVTNNATLNNQSKLSAKIMELNTHTLNTSKLSQAKNITKTEMLN 540
QY 370 GTAPIYNYLPTSLYLOK-----NOTLA-----TIKAKENO-LSQIAPNN-- 409
DB 541 GEA-----SLVAKEDINAKIDKTNNGTITAGIZANITTKALENDNALILAOULN 591
QY 410 -----YI-----PSKNLAPIALAADOSS-----TPITMANTNOFLEKTKOL 448
DB 592 FTYNGSHYVNGDIVSKDAIYFSSNSDFTSNGSKLVDAONNLTVANNENITQSSSEI 651
QY 449 RLDTDOYVYNT--ATYNEEN-GRVAVDTGSNMSEVLPQIOETTARILFNGKDLNVER- 503
DB 652 -----LHGNTVLNAKGNFTNSGNLTJTKRELNTSINSEFI--NAGMLTJGKLEVHSNT 702
QY 504 -----RIAAVNSDPLETTKPDMLT--KEALKIAGF--FNEPNGILOYGADITEF 550
DB 703 TVANDGKLVSTENLNTISKTDFTNNGTLLGLEALKIASGNGFTVNASGSLA--SNKSLDIY 761
QY 551 DFNF--DOOTSNIKINOLAEINATNIYVLDK--IKLNAMNILLRDRFRHYDRNNIAGA 607

```

```

Db 762 GNNFTNNGTIESYKS-----LNTNNTFTFNNTATIKSYGVNLNTSOG-NFTNDSNGTYMSH 816
QY 608 D-ESYVKEAHEVYNSST-----BGLL---NIDKDIKILSGYIVEIEDTEGLKEVIND 658
Db 817 DLNLTISOAN--LNNKLLAGGGLNLTAKGNITNDSNSTALIAVLSNNDIN-----LNA 869
QY 659 RYDMLNSSL-RDCKTFIDPKYNDKLPYLSNPNKVVAVYAVTKENTINPSENGDTS 717
Db 870 NKKYVNIIGEYISOGNISVEAKLHNDVKL-----SGNITTTTK-----SGNATVK 915
QY 718 TNGI 721
Db 916 TNSI 919

```

Search completed: October 4, 2002, 14:40:05
 Job time: 280 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 4, 2002, 14:34:15 ; Search time 14.66 Seconds
(without alignments) 1226.278 Million cell updates/sec

Title: US-09-848-909-21

Perfect score: 1 EVKQENRLNLSSESSGGL.....STNGIKKILFSSKKEIEZ 736

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued Patents, AA:*

- 1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
- 4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
- 5: /cgn2_6/ptodata/2/1aa/PCPTS.COMB.pep.*
- 6: /cgn2_6/ptodata/2/1aa/Backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3774	99.9	735	1	US-08-021-601-4
2	3774	99.9	735	1	US-08-021-601-4
3	3774	99.9	735	1	US-08-021-601-4
4	3729.5	98.7	903	1	US-08-021-601-4
5	3729.5	98.7	903	1	US-08-021-601-4
6	3729.5	98.7	903	1	US-08-021-601-4
7	3597	95.2	719	1	US-08-082-843B-12
8	3597	95.2	719	1	US-08-082-843B-12
9	3597	95.2	719	1	US-08-082-843B-12
10	3597	95.2	719	1	US-08-082-843B-12
11	3597	95.2	719	1	US-08-082-843B-12
12	3597	95.2	719	1	US-08-082-843B-12
13	3597	95.2	719	1	US-08-082-843B-12
14	3597	95.2	719	1	US-08-082-843B-12
15	3597	95.2	719	1	US-08-082-843B-12
16	3597	95.2	719	1	US-08-082-843B-12
17	3597	95.2	719	1	US-08-082-843B-12
18	3597	95.2	719	1	US-08-082-843B-12
19	3597	95.2	719	1	US-08-082-843B-12
20	3597	95.2	719	1	US-08-082-843B-12
21	3597	95.2	719	1	US-08-082-843B-12
22	3597	95.2	719	1	US-08-082-843B-12
23	3597	95.2	719	1	US-08-082-843B-12
24	3597	95.2	719	1	US-08-082-843B-12
25	3597	95.2	719	1	US-08-082-843B-12
26	3597	95.2	719	1	US-08-082-843B-12
27	3597	95.2	719	1	US-08-082-843B-12

28	789	20.9	852	2	US-08-463-483A-36	Sequence 36, Appl
29	789	20.9	852	2	US-08-471-046A-36	Sequence 36, Appl
30	789	20.9	852	2	US-08-470-566B-36	Sequence 36, Appl
31	789	20.9	852	2	US-08-469-334-36	Sequence 36, Appl
32	789	20.9	852	2	US-09-300-529-36	Sequence 36, Appl
33	789	20.9	1338	1	US-08-471-043A-50	Sequence 50, Appl
34	789	20.9	1338	1	US-08-471-044-50	Sequence 50, Appl
35	789	20.9	1338	2	US-08-463-483A-50	Sequence 50, Appl
36	789	20.9	1338	2	US-08-470-566B-50	Sequence 50, Appl
37	789	20.9	1338	2	US-08-470-566B-50	Sequence 50, Appl
38	789	20.9	1338	2	US-08-469-334-50	Sequence 50, Appl
39	789	20.9	1338	3	US-09-300-529-50	Sequence 50, Appl
40	775.5	20.5	834	4	US-08-471-043A-21	Sequence 21, Appl
41	756.5	20.0	834	1	US-08-471-044-21	Sequence 21, Appl
42	756.5	20.0	834	2	US-08-463-483A-21	Sequence 21, Appl
43	756.5	20.0	834	2	US-08-470-566B-21	Sequence 21, Appl
44	756.5	20.0	834	2	US-08-471-046A-21	Sequence 21, Appl
45	756.5	20.0	834	2	US-08-470-566B-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-08-021-601-4
Sequence 4, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Kimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Alora, Naveen
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Protein Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021.601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.057
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 735 amino acids
TYPE: AMINO ACID
MOLECULE TYPE: protein
US-08-021-601-4
Query Match 99.9%; Score 3774; DB 1; Length 735;
Best Local Similarity 100.0%; Pred. No. 1.9e-258;
Matches 735; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1. EVKQENRLNLSSESSGGLGYFDLNPQAPMVYTSSTGDLSPSELENIPEENQVF 60

```

Db      1 EVQOERRLNLSSESSSGGLGYFSDLFQAPWVYSSSTGDLSPSELENIPESENOYF 60
Qy      61 QSAIWSGFIKAKKDEYTPATSDADNHVTMMVDDQEVINKASNSKRLKLGRLYQIKIY 120
Db      61 QSAIWSGFIKAKKDEYTPATSDADNHVTMMVDDQEVINKASNSKRLKLGRLYQIKIY 120
Qy      121 QRENPEKGLDFKLYMTDSQNKKEVYSSDNLOJPELKOKSSNSKRRKSTASGPTYPDRDN 180
Db      121 QRENPEKGLDFKLYMTDSQNKKEVYSSDNLOJPELKOKSSNSKRRKSTASGPTYPDRDN 180
Qy      181 DGIPDSLEVEGYTYDVAKNRKTFSPWISNIHEKKGLTKKSSPEKMTASDPYSEKVT 240
Db      181 DGIPDSLEVEGYTYDVAKNRKTFSPWISNIHEKKGLTKKSSPEKMTASDPYSEKVT 240
Qy      241 GRIDKVVSPKARHPLVAAYPIVHVMENIILSKNEQSTQNTDSEFTISKNTSRTHT 300
Db      241 GRIDKVVSPKARHPLVAAYPIVHVMENIILSKNEQSTQNTDSEFTISKNTSRTHT 300
Qy      301 SEVHGAELVHASFDFDIGSVSAGFSNSNSTVAIDHSLSLAGEETVAETMGLTADTARL 360
Db      301 SEVHGAELVHASFDFDIGSVSAGFSNSNSTVAIDHSLSLAGEETVAETMGLTADTARL 360
Qy      361 NANIRVNTGAPLYNVLPFTSLVJGKNQTLATIKAKENQSLQILAPNNYPSKMLAPIA 420
Db      361 NANIRVNTGAPLYNVLPFTSLVJGKNQTLATIKAKENQSLQILAPNNYPSKMLAPIA 420
Qy      421 LNAODFSSPTITANNYNOFLELEKTKOLRLDTDOVYGNIAFYENENGRVAVDTGSNMSEV 480
Db      421 LNAODFSSPTITANNYNOFLELEKTKOLRLDTDOVYGNIAFYENENGRVAVDTGSNMSEV 480
Qy      481 LPOIOTETARIIFNGKDLNIVERRIAAVNSDDPLETTKPDMTLKEALKIAGFENPNL 540
Db      481 LPOIOTETARIIFNGKDLNIVERRIAAVNSDDPLETTKPDMTLKEALKIAGFENPNL 540
Qy      541 OYQGRITPEPNDQOOTSQNKIKNOLAELANTNITYVLDKIKLAKANNIILIRKREPHDR 600
Db      541 OYQGRITPEPNDQOOTSQNKIKNOLAELANTNITYVLDKIKLAKANNIILIRKREPHDR 600
Qy      601 NNIAVGADESVYKKAHREYINNSTBGLLNDIDDKILISGTYEIEDTGLKEVIYNDRY 660
Db      601 NNIAVGADESVYKKAHREYINNSTBGLLNDIDDKILISGTYEIEDTGLKEVIYNDRY 660
Qy      661 DMLNISLSLQDGKTFIDFKYNDKLPYISNPNKYVNAVYATKENTLIINSENGDTSTNG 720
Db      661 DMLNISLSLQDGKTFIDFKYNDKLPYISNPNKYVNAVYATKENTLIINSENGDTSTNG 720
Qy      721 IKKILIFSCKGEYEG 735
Db      721 IKKILIFSCKGEYEG 735

RESULT 2
US-08-082-849B-4
; Sequence 4, Application US/08082849B
; Patent No. 5677274
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Kimpel, Kurt R.
; APPLICANT: Arota, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:

```

```

; MEDIAN TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/082,849B
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/021,601
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-161-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 735 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-082-849B-4

Query Match 99.9%; Score 3774; DB 1; Length 735;
Best Local Similarity 100.0%; Pred. No. 1.9e-258;
Matches 735; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 EVQOERRLNLSSESSSGGLGYFSDLFQAPWVYSSSTGDLSPSELENIPESENOYF 60
1 EVQOERRLNLSSESSSGGLGYFSDLFQAPWVYSSSTGDLSPSELENIPESENOYF 60
61 QSAIWSGFIKAKKDEYTPATSDADNHVTMMVDDQEVINKASNSKRLKLGRLYQIKIY 120
61 QSAIWSGFIKAKKDEYTPATSDADNHVTMMVDDQEVINKASNSKRLKLGRLYQIKIY 120
121 QRENPEKGLDFKLYMTDSQNKKEVYSSDNLOJPELKOKSSNSKRRKSTASGPTYPDRDN 180
121 QRENPEKGLDFKLYMTDSQNKKEVYSSDNLOJPELKOKSSNSKRRKSTASGPTYPDRDN 180
181 DGIPDSLEVEGYTYDVAKNRKTFSPWISNIHEKKGLTKKSSPEKMTASDPYSEKVT 240
181 DGIPDSLEVEGYTYDVAKNRKTFSPWISNIHEKKGLTKKSSPEKMTASDPYSEKVT 240
241 GRIDKVVSPKARHPLVAAYPIVHVMENIILSKNEQSTQNTDSEFTISKNTSRTHT 300
241 GRIDKVVSPKARHPLVAAYPIVHVMENIILSKNEQSTQNTDSEFTISKNTSRTHT 300
301 SEVHGAELVHASFDFDIGSVSAGFSNSNSTVAIDHSLSLAGEETVAETMGLTADTARL 360
301 SEVHGAELVHASFDFDIGSVSAGFSNSNSTVAIDHSLSLAGEETVAETMGLTADTARL 360
361 NANIRVNTGAPLYNVLPFTSLVJGKNQTLATIKAKENQSLQILAPNNYPSKMLAPIA 420
361 NANIRVNTGAPLYNVLPFTSLVJGKNQTLATIKAKENQSLQILAPNNYPSKMLAPIA 420
421 LNAODFSSPTITANNYNOFLELEKTKOLRLDTDOVYGNIAFYENENGRVAVDTGSNMSEV 480
421 LNAODFSSPTITANNYNOFLELEKTKOLRLDTDOVYGNIAFYENENGRVAVDTGSNMSEV 480
481 LPOIOTETARIIFNGKDLNIVERRIAAVNSDDPLETTKPDMTLKEALKIAGFENPNL 540
481 LPOIOTETARIIFNGKDLNIVERRIAAVNSDDPLETTKPDMTLKEALKIAGFENPNL 540
541 OYQGRITPEPNDQOOTSQNKIKNOLAELANTNITYVLDKIKLAKANNIILIRKREPHDR 600
541 OYQGRITPEPNDQOOTSQNKIKNOLAELANTNITYVLDKIKLAKANNIILIRKREPHDR 600
601 NNIAVGADESVYKKAHREYINNSTBGLLNDIDDKILISGTYEIEDTGLKEVIYNDRY 660
601 NNIAVGADESVYKKAHREYINNSTBGLLNDIDDKILISGTYEIEDTGLKEVIYNDRY 660

```

DB 601 NN1AVGADSVYKKAHREYVINSSTEGLLNIDKDIRKILSGYIWELEDEGLKEVINDRY 660
QY 661 DMLN1SLRQDKTEFIDFKKYNKDLPLYISNPNKYKVAATKENTLINSNGSTNG 720
DB 661 DMLN1SLRQDKTEFIDFKKYNKDLPLYISNPNKYKVAATKENTLINSNGSTNG 720
QY 721 IKKILFSKKGYEIG 735
DB 721 IKKILFSKKGYEIG 735

RESULT 3
PCT-US94-01624-4
Sequence 4, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppia, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Arota, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
RELATED METHODS
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESSES:
ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE and CREW
STREET: Stewart Street Tower, 20th Floor, One Market
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 735 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-4

Query Match 99.9% Score 3774: DB 5: length 735;
Best Local Similarity 100.0% Pctd No. 1.9e-258;
Matches 735: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EVKQENRLLNLSSESSQGLIYFSDLNFOAPVVTSTTGDISSPSSLENINSEKQTF 60
DB 1 EVKQENRLLNLSSESSQGLIYFSDLNFOAPVVTSTTGDISSPSSLENINSEKQTF 60
QY 61 GSAINSGFIYKKSDEYTFATSDNHNVTMVDQEVINKASNSNKTRELKGLQIKIOT 120
DB 61 GSAINSGFIYKKSDEYTFATSDNHNVTMVDQEVINKASNSNKTRELKGLQIKIOT 120
QY 121 QRENTEGGLDFKLYWTDSONKKEVSSDNLQPELKOKSSNSRKRRTSAGPTVPDRDN 180
DB 121 QRENTEGGLDFKLYWTDSONKKEVSSDNLQPELKOKSSNSRKRRTSAGPTVPDRDN 180
QY 181 DGIPSLSEGEYTVVKNKRTFLSPWISNIHKKGLTKYKSPKMSYSDPYSDPEKYV 240

DB 181 DGIPSLSEGEYTVVKNKRTFLSPWISNIHKKGLTKYKSPKMSYSDPYSDPEKYV 240
QY 241 GRIDKNVSPPEARHPLVAAPYIVHVDMENTILSKNEQOSTONTDSETRTISKNTSTRT 300
DB 241 GRIDKNVSPPEARHPLVAAPYIVHVDMENTILSKNEQOSTONTDSETRTISKNTSTRT 300
QY 301 SEVHGNAVHASFEDIGGSVSAGFSNSNSTVAIDHSLSLAGEYTNATMGINTADTARL 360
DB 301 SEVHGNAVHASFEDIGGSVSAGFSNSNSTVAIDHSLSLAGEYTNATMGINTADTARL 360
QY 361 NANIRVYNTGAPTYNVLPTSLVYGRKQTLATIKAKENOLSOILANNNYPSKMLAPIA 420
DB 361 NANIRVYNTGAPTYNVLPTSLVYGRKQTLATIKAKENOLSOILANNNYPSKMLAPIA 420
QY 421 LNKODPSTPTNNYNOLELEKTKOLRDTDOVGNATYVNSPENGNAVDTGSNMGSEV 480
DB 421 LNKODPSTPTNNYNOLELEKTKOLRDTDOVGNATYVNSPENGNAVDTGSNMGSEV 480
QY 481 LPOIOTTRKRIIFNGKDLNVERRIAAVNSPDELTTRKPDMTIKAKLAFGNPNNGNL 540
DB 481 LPOIOTTRKRIIFNGKDLNVERRIAAVNSPDELTTRKPDMTIKAKLAFGNPNNGNL 540
QY 541 QYQKDIETEDFNEDQOSTONIKNOLAEKNTIYVLDKIKLWKNLILRDKRPHYDR 600
DB 541 QYQKDIETEDFNEDQOSTONIKNOLAEKNTIYVLDKIKLWKNLILRDKRPHYDR 600
QY 601 NN1AVGADSVYKKAHREYVINSSTEGLLNIDKDIRKILSGYIWELEDEGLKEVINDRY 660
DB 601 NN1AVGADSVYKKAHREYVINSSTEGLLNIDKDIRKILSGYIWELEDEGLKEVINDRY 660
QY 661 DMLN1SLRQDKTEFIDFKKYNKDLPLYISNPNKYKVAATKENTLINSNGSTNG 720
DB 661 DMLN1SLRQDKTEFIDFKKYNKDLPLYISNPNKYKVAATKENTLINSNGSTNG 720
QY 721 IKKILFSKKGYEIG 735
DB 721 IKKILFSKKGYEIG 735

RESULT 4
US-08-021-601-12
Sequence 12, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:
APPLICANT: Leppia, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Arota, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
RELATED METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Owendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.057

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 404/688-0770
 TELEFAX: 404/688-9880
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 903 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-021-601-12

Query Match 98.7%; Score 3729.5; DB 1; Length 903;
 Best Local Similarity 99.2%; Pred. No. 3; se-255;
 Matches 728; Conservative 1; Mismatches 2; Indels 3; Gaps 1;

```

QY 1 EVKQENRLNSESSESSGGLGYFSDLNFOAPMVVTSSTTGDLSPSSSELENIPSENOYF 60
DB 1 EVKQENRLNSESSESSGGLGYFSDLNFOAPMVVTSSTTGDLSPSSSELENIPSENOYF 60
QY 61 QSAIWSGFIKVKKSDEYTFATSADNHVYTMVWDQEVYINKASNSKIRLEKGRLYOIKIOY 120
DB 61 QSAIWSGFIKVKKSDEYTFATSADNHVYTMVWDQEVYINKASNSKIRLEKGRLYOIKIOY 120
QY 121 QRENPTKGLDFKLYMTDSQNKKEVYSSDNLOLPELKOKSSNSRKRKSTAGPTVPDRDN 180
DB 121 QRENPTKGLDFKLYMTDSQNKKEVYSSDNLOLPELKOKSSNSRKRKSTAGPTVPDRDN 180
QY 181 DGIPDLSLEVEGYTYDVKNKRTFLSPWISNIHEKKGLTKYKSSPEKWSADPYDPEKYT 240
DB 181 DGIPDLSLEVEGYTYDVKNKRTFLSPWISNIHEKKGLTKYKSSPEKWSADPYDPEKYT 240
QY 241 GRIDKNVSPPEARHPLVAAPYIVHVMENIILSKNEQOSTONTDEFTRTSKNTSRTHT 300
DB 241 GRIDKNVSPPEARHPLVAAPYIVHVMENIILSKNEQOSTONTDEFTRTSKNTSRTHT 300
QY 301 SEVHGNAEVHASFPDIGGSVAGFSNSNSTVAIDHSLSLAGERTVAETMGNTADTARL 360
DB 301 SEVHGNAEVHASFPDIGGSVAGFSNSNSTVAIDHSLSLAGERTVAETMGNTADTARL 360
QY 361 NANIRVYNTGTAPIYVNLPTTSLVYGNKOTLATIKAKENQSLQILAPNNYYPKNIAPIA 420
DB 361 NANIRVYNTGTAPIYVNLPTTSLVYGNKOTLATIKAKENQSLQILAPNNYYPKNIAPIA 420
QY 421 LNNODFSSPTITNNYNOFLELEKTKOLRLDPOYGNATATNPNKRVYVDPGSMSEY 480
DB 421 LNNODFSSPTITNNYNOFLELEKTKOLRLDPOYGNATATNPNKRVYVDPGSMSEY 480
QY 481 LPOIETFTARIIFNGKDLNVERRIAANVPSDPLETTKEDMTLKAALKAFGEFNEPNGNL 540
DB 481 LPOIETFTARIIFNGKDLNVERRIAANVPSDPLETTKEDMTLKAALKAFGEFNEPNGNL 540
QY 541 OYQKQITTEFPDNDQOSTONIKNOLAELANVTYVLDKIKLAKNNMILLRDKRFHYDR 600
DB 541 OYQKQITTEFPDNDQOSTONIKNOLAELANVTYVLDKIKLAKNNMILLRDKRFHYDR 600
QY 601 NNIAVADESVYKKAHREVIINSSTBGLLNDKDKIRKILSGYVEIEDTEGLEKVIANDRY 660
DB 601 NNIAVADESVYKKAHREVIINSSTBGLLNDKDKIRKILSGYVEIEDTEGLEKVIANDRY 660
QY 661 DMLNSSLRODGKTFIDFKKYNKDKLPLIISNPNKVVYAVATKENTJINSNGDSTISNG 720
DB 661 DMLNSSLRODGKTFIDFKKYNKDKLPLIISNPNKVVYAVATKENTJINSNGDSTISNG 720
QY 721 IKRIL--IFSKKG 731
DB 721 IKRILKKVYLGKG 734

```

GENERAL INFORMATION:
 APPLICANT: Leppla, Stephen H.
 APPLICANT: Klimpel, Kurt R.
 APPLICANT: Aroca, Naveen
 APPLICANT: Singh, Yogendra
 APPLICANT: Nichols, Peter J.
 TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
 TITLE OF INVENTION: Related Methods
 NUMBER OF SEQUENCES: 35
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: PC-DOS/MS-DOS
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/082,849B
 FILING DATE: 25-JUN-1993
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/021,601
 FILING DATE: 12-FEB-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Weber, Kenneth A.
 REGISTRATION NUMBER: 31,677
 REFERENCE/DOCKET NUMBER: 15280-161-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0300
 TELEFAX: (415) 576-0200
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 903 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-082-849B-12

Query Match 98.7%; Score 3729.5; DB 1; Length 903;
 Best Local Similarity 99.2%; Pred. No. 3; se-255;
 Matches 728; Conservative 1; Mismatches 2; Indels 3; Gaps 1;

```

QY 1 EVKQENRLNSESSESSGGLGYFSDLNFOAPMVVTSSTTGDLSPSSSELENIPSENOYF 60
DB 1 EVKQENRLNSESSESSGGLGYFSDLNFOAPMVVTSSTTGDLSPSSSELENIPSENOYF 60
QY 61 QSAIWSGFIKVKKSDEYTFATSADNHVYTMVWDQEVYINKASNSKIRLEKGRLYOIKIOY 120
DB 61 QSAIWSGFIKVKKSDEYTFATSADNHVYTMVWDQEVYINKASNSKIRLEKGRLYOIKIOY 120
QY 121 QRENPTKGLDFKLYMTDSQNKKEVYSSDNLOLPELKOKSSNSRKRKSTAGPTVPDRDN 180
DB 121 QRENPTKGLDFKLYMTDSQNKKEVYSSDNLOLPELKOKSSNSRKRKSTAGPTVPDRDN 180
QY 181 DGIPDLSLEVEGYTYDVKNKRTFLSPWISNIHEKKGLTKYKSSPEKWSADPYDPEKYT 240
DB 181 DGIPDLSLEVEGYTYDVKNKRTFLSPWISNIHEKKGLTKYKSSPEKWSADPYDPEKYT 240
QY 241 GRIDKNVSPPEARHPLVAAPYIVHVMENIILSKNEQOSTONTDEFTRTSKNTSRTHT 300
DB 241 GRIDKNVSPPEARHPLVAAPYIVHVMENIILSKNEQOSTONTDEFTRTSKNTSRTHT 300
QY 301 SEVHGNAEVHASFPDIGGSVAGFSNSNSTVAIDHSLSLAGERTVAETMGNTADTARL 360
DB 301 SEVHGNAEVHASFPDIGGSVAGFSNSNSTVAIDHSLSLAGERTVAETMGNTADTARL 360
QY 361 NANIRVYNTGTAPIYVNLPTTSLVYGNKOTLATIKAKENQSLQILAPNNYYPKNIAPIA 420

```



```

|||||
Db 361 NANIRVNTGAPLYVNLPTSTLVGKNOTLATIKAKENOLSOILAPNNYPSKMLAFIA 420
Oy 421 LNAODFSSPTITNNYNOFLELEKTKOLRLDQVGNATATYNNENGRVAVTGSNMSEV 480
Db 421 LNAODFSSPTITNNYNOFLELEKTKOLRLDQVGNATATYNNENGRVAVTGSNMSEV 480
Oy 481 LPQIETARTIIFNGKDLNVERRIAANVPSDPLETTKPDWTLKALKAFGNPNL 540
Db 481 LPQIETARTIIFNGKDLNVERRIAANVPSDPLETTKPDWTLKALKAFGNPNL 540
Oy 541 OYQGDITEFPNPDQTSNKNKLAELNATNTYVLDKIKLAKNNLIDKRFHNR 600
Db 541 OYQGDITEFPNPDQTSNKNKLAELNATNTYVLDKIKLAKNNLIDKRFHNR 600
Oy 601 NNANAGADESVYKAAHREYVNSTGGLLNIDKDKILSGYVEIEDTGLKEVINDRY 660
Db 601 NNANAGADESVYKAAHREYVNSTGGLLNIDKDKILSGYVEIEDTGLKEVINDRY 660
|||||
DQKTFIDFKYNDKLPYISNPNYKVNVAATKENTLINSENGDSTNG 720
DQKTFIDFKYNDKLPYISNPNYKVNVAATKENTLINSENGDSTNG 720
SRKG 731
GKKG 734
GKKG 734
|||||
cation PC/TUS9401624
ON:
pla, Stephen H.
mpel, Kurt R.
ra, Naveen
gh, Yogendra
hols, Peter J.
|||||
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESS: TOWNSEND AND TOWNSEND KHOURIE AND CREW
STREET: Stewart Street Tower, 20th Floor, One Market
STREET: Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-12

```

```

Query Match 98.7%; Score 3729.5; DB 5; Length 903;
Best Local Similarity 99.2%; Pred. No. 3, 5e-255;
Matches 728; Conservative 1; Mismatches 2; Indels 3; Gaps 1;
Oy 1 EYKQENRLNNESSSSOGILGYFSDLPQAPVWVSSSTGDLNPSSELENIPEENGYF 60
Db 1 EYKQENRLNNESSSSOGILGYFSDLPQAPVWVSSSTGDLNPSSELENIPEENGYF 60
Oy 61 QSAIISGFIRYKKSDEYTFATSDNHTVMWDQEVYKASNSNKLRLKGLYQIKYQ 120
Db 61 QSAIISGFIRYKKSDEYTFATSDNHTVMWDQEVYKASNSNKLRLKGLYQIKYQ 120
Oy 121 QRENPTKGLDFKLTWDSQNKKEYISSDNLQPLKOKSSNKKRSTASAGTPPDDN 180
Db 121 QRENPTKGLDFKLTWDSQNKKEYISSDNLQPLKOKSSNKKRSTASAGTPPDDN 180
Oy 181 DGIPLSLEBEGTYVDKAKRFTLSFWISNHEKKGLTKYKSSPKNASTASDPYSPFVY 240
Db 181 DGIPLSLEBEGTYVDKAKRFTLSFWISNHEKKGLTKYKSSPKNASTASDPYSPFVY 240
Oy 241 GRIDKNSPEARHPLVAAPVIVHDMENILSKNEQSTONTDSERTISKNTSTSRHT 300
Db 241 GRIDKNSPEARHPLVAAPVIVHDMENILSKNEQSTONTDSERTISKNTSTSRHT 300
Oy 301 SEVHGNAEVHASFEDIGGSVSAFSSNSSTVAIDHSLSLAGEEVAETMGINTADTARL 360
Db 301 SEVHGNAEVHASFEDIGGSVSAFSSNSSTVAIDHSLSLAGEEVAETMGINTADTARL 360
Oy 361 NANIRVNTGAPLYVNLPTSTLVGKNOTLATIKAKENOLSOILAPNNYPSKMLAFIA 420
Db 361 NANIRVNTGAPLYVNLPTSTLVGKNOTLATIKAKENOLSOILAPNNYPSKMLAFIA 420
Oy 421 LNAODFSSPTITNNYNOFLELEKTKOLRLDQVGNATATYNNENGRVAVTGSNMSEV 480
Db 421 LNAODFSSPTITNNYNOFLELEKTKOLRLDQVGNATATYNNENGRVAVTGSNMSEV 480
Oy 481 LPQIETARTIIFNGKDLNVERRIAANVPSDPLETTKPDWTLKALKAFGNPNL 540
Db 481 LPQIETARTIIFNGKDLNVERRIAANVPSDPLETTKPDWTLKALKAFGNPNL 540
Oy 541 OYQGDITEFPNPDQTSNKNKLAELNATNTYVLDKIKLAKNNLIDKRFHNR 600
Db 541 OYQGDITEFPNPDQTSNKNKLAELNATNTYVLDKIKLAKNNLIDKRFHNR 600
Oy 601 NNANAGADESVYKAAHREYVNSTGGLLNIDKDKILSGYVEIEDTGLKEVINDRY 660
Db 601 NNANAGADESVYKAAHREYVNSTGGLLNIDKDKILSGYVEIEDTGLKEVINDRY 660
Oy 661 DMLNSSLRQDKTFIDFKYNDKLPYISNPNYKVNVAATKENTLINSENGDSTNG 720
Db 661 DMLNSSLRQDKTFIDFKYNDKLPYISNPNYKVNVAATKENTLINSENGDSTNG 720
Oy 721 IKKIL--IFSKG 731
Db 721 IKKILKKYVIGKKG 734
|||||
RESULT 7
US-08-849B-31
Sequence 31, Application US/08082849B
Patent No. 567274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klompel, Kurt R.
APPLICANT: Aroa, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor

```

```

CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0300
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 719 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-31

```

```

Query Match          95.2%; Score 3597; DB 1; Length 719;
Best Local Similarity 95.5%; Pred. No. 5,7e-246;
Matches 706; Conservative 4; Mismatches 5; Indels 24; Gaps 2;

QY 1 EVKQENRLNLESESSSSQGLGYFSDLNFOAPMVVYTSSTGDLSPSSSELENIPSENOYF 60
DB 1 EVKQENRLNLESESSSSQGLGYFSDLNFOAPMVVYTSSTGDLSPSSSELENIPSENOYF 60
QY 61 QSAIWSGFIRKVKSDSEYFATSADNHVTMMVWDDQEVINKASNSNKRILEKGRLYQIKIY 120
DB 61 QSAIWSGFIRKVKSDSEYFATSADNHVTMMVWDDQEVINKASNSNKRILEKGRLYQIKIY 120
QY 121 QRENPEKGLDEKLYWTDSONKKEVYSSDNLOLPELKOKSSNS-----RKRSTYSAQPTVP 176
DB 121 QRENPEKGLDEKLYWTDSONKKEVYSSDNLOLPELKOKSSNS-----RKRSTYSAQPTVP 180
QY 177 DRONDGIPDSLEVEGYVDVKNKRFFELSPWISNIHEKGLTRYKSSPEKMSSTASPYGDF 236
DB 177 DRONDGIPDSLEVEGYVDVKNKRFFELSPWISNIHEKGLTRYKSSPEKMSSTASPYGDF 240
QY 181 DRONDGIPDSLEVEGYVDVKNKRFFELSPWISNIHEKGLTRYKSSPEKMSSTASPYGDF 240
DB 181 DRONDGIPDSLEVEGYVDVKNKRFFELSPWISNIHEKGLTRYKSSPEKMSSTASPYGDF 240
QY 237 EKYVGRIDKNVSEPARHPLVAAPIVHVDMENITLSKNEIDOSTOWTSETRTSKNTSTS 296
DB 237 EKYVGRIDKNVSEPARHPLVAAPIVHVDMENITLSKNEIDOSTOWTSETRTSKNTSTS 300
QY 241 EKYVGRIDKNVSEPARHPLVAAPIVHVDMENITLSKNEIDOSTOWTSETRTSKNTSTS 300
DB 241 EKYVGRIDKNVSEPARHPLVAAPIVHVDMENITLSKNEIDOSTOWTSETRTSKNTSTS 300
QY 297 RTHHSVHNAVYASFPDIGSVSAGFSNSNSTVAIDHSISLAGRTWAEFMGLNTRAD 356
DB 297 RTHHSVHNAVYASFPDIGSVSAGFSNSNSTVAIDHSISLAGRTWAEFMGLNTRAD 360
QY 301 RTHHSVHNAVYASFPDIGSVSAGFSNSNSTVAIDHSISLAGRTWAEFMGLNTRAD 360
DB 301 RTHHSVHNAVYASFPDIGSVSAGFSNSNSTVAIDHSISLAGRTWAEFMGLNTRAD 360
QY 357 TARLANNTYVVTGAPRIYVLPITSLVIGKQTLATIRAKNOLSOILAPNNYPSKNL 416
DB 357 TARLANNTYVVTGAPRIYVLPITSLVIGKQTLATIRAKNOLSOILAPNNYPSKNL 420
QY 417 APALANADDESSPTITMANTNOFLELEKTKOLRLDDOYVGNATVYFENGVRVDTGSN 476
DB 417 APALANADDESSPTITMANTNOFLELEKTKOLRLDDOYVGNATVYFENGVRVDTGSN 480
QY 477 WSEVLQIOETTARLIINGKDLNVERRIAANVNSDPLETTPDMATKEALKIAGFNEP 536
DB 477 WSEVLQIOETTARLIINGKDLNVERRIAANVNSDPLETTPDMATKEALKIAGFNEP 540
QY 537 NGNIQYOGDIIEFDENFDQOTSQNIKQNLAEIANTYTVLDKIKINAMNILLRDRF 596
DB 537 NGNIQYOGDIIEFDENFDQOTSQNIKQNLAEIANTYTVLDKIKINAMNILLRDRF 596

```

```

DB 521 NGNIQYOGDIIEFDENFDQOTSQNIKQNLAEIANTYTVLDKIKINAMNILLRDRF 580
QY 597 HYDRNNINAGADESVYKAEAREVINSSTEGLLINDKIDKIRKILSGTYVEDETGLEKEV 656
DB 581 HYDRNNINAGADESVYKAEAREVINSSTEGLLINDKIDKIRKILSGTYVEDETGLEKEV 640
QY 657 NDRYDMLNISLQKDKTEFIDKRYNDKLPYISNPNKVVYAVFKENTIIINSENGDT 716
DB 641 NDRYDMLNISLQKDKTEFIDKRYNDKLPYISNPNKVVYAVFKENTIIINSENGDT 700
QY 717 STNGIKRILIFSKKGYEIG 735
DB 701 STNGIKRILIFSKKGYEIG 719

```

```

RESULT 8
PCT-US94-01624-31
Sequence 31, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Arota, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESS: TOWNSEND AND TOWNSEND KHOURIE AND GREEN
STREET: Steuart Street Tower, 20th floor, One Market
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 719 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-31

```

```

Query Match          95.2%; Score 3597; DB 5; Length 719;
Best Local Similarity 95.5%; Pred. No. 5,7e-246;
Matches 706; Conservative 4; Mismatches 5; Indels 24; Gaps 2;

QY 1 EVKQENRLNLESESSSSQGLGYFSDLNFOAPMVVYTSSTGDLSPSSSELENIPSENOYF 60
DB 1 EVKQENRLNLESESSSSQGLGYFSDLNFOAPMVVYTSSTGDLSPSSSELENIPSENOYF 60
QY 61 QSAIWSGFIRKVKSDSEYFATSADNHVTMMVWDDQEVINKASNSNKRILEKGRLYQIKIY 120
DB 61 QSAIWSGFIRKVKSDSEYFATSADNHVTMMVWDDQEVINKASNSNKRILEKGRLYQIKIY 120

```

Fri Oct 4 14:31:03 2002

us-09-848-909-21.ra1

Page 7

121 QRENPTEKGLDFKLTWDSQNKKEVYSSDNLOPELAKOSKSSNS-----RKRSSTGAPTV 176
121 QRENPTEKGLDFKLTWDSQNKKEVYSSDNLOPELAKOSKSSNTATIMOGCNLGGPVP 180
177 DRONDIGPDLSEVEGYTVDKNKRFTFLSPWISNHEKGLTKYKSPKSTSPDYSD 236
181 DRONDIGPDLSEVEGYTVDKNKRFTFLSPWISNHEKGLTKYKSPKSTSPDYSD 240
227 EKVGRIDKNVSPPEARHPLVAAPPIVYVDMENILLSKNEDOSTONTDETIRISKTST 296
241 EKVGRIDKNVSPPEARHPLVAAPPIVYVDMENILLSKNEDOSTONTDETIRISKTST 300
297 RHTSEVHGAAVHAFPDIGSVAGSAGFNSNSTVAIDHSLSLAGEPTMAETMGLNTAD 356
301 RHTSEVHGAAVHAFPDIGSVAGSAGFNSNSTVAIDHSLSLAGEPTMAETMGLNTAD 360
357 TARKLANIRVNTGTAPLYNVLPTSLVYKGNQTLATIKAKENQLSOILAPNNYPSKML 416
361 TARKLANIRVNTGTAPLYNVLPTSLVYKGNQTLATIKAKENQLSOILAPNNYPSKML 420
417 APIALNADDFSTPTNNYKQFLEKTKQKLDQOYGNATYFENGVRVDTGSN 476
421 APIALNADDFSTPTNNYKQFLEKTKQKLDQOYGNATYFENGVRVDTGSN 460
477 WSEVLPJOETRTARITENGKOLANVERRIAAVNSPDLPTTKPDMTKFAKIAFGNEP 536
461 WSEVLPJOETRTARITENGKOLANVERRIAAVNSPDLPTTKPDMTKFAKIAFGNEP 520
537 NGNLQYOGKDTTEPFDPOOTSNIKOLAELNATNTYVLDKIKLNKNNILRDKRF 596
521 NGNLQYOGKDTTEPFDPOOTSNIKOLAELNATNTYVLDKIKLNKNNILRDKRF 580
597 HYDRNINAVGADESUYKAEHREYINSSTEGILLNIDKIRKILSGYVEIEDEBLKRYI 656
581 HYDRNINAVGADESUYKAEHREYINSSTEGILLNIDKIRKILSGYVEIEDEBLKRYI 640
657 NDRYDLNITSLRODKTFIDEKYNKDLPLYSNPNRYNVAATKENTINSENGDT 716
641 NDRYDLNITSLRODKTFIDEKYNKDLPLYSNPNRYNVAATKENTINSENGDT 700
717 STNGIKKILIFSKGYEIG 735
701 STNGIKKILIFSKGYEIG 719
RESULT 9
US-09-273-839A-8
Sequence 8, Application US/09273839A
GENERAL INFORMATION:
APPLICANT: Clitino, Nick M
APPLICANT: Jackson, Bruce E
TITLE OF INVENTION: Disruption of Anthrax Toxin Binding to Cell Surface
FILE REFERENCE: S-89/662
CURRENT APPLICATION NUMBER: US/09/273,839A
CURRENT FILING DATE: 1999-03-22
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8
LENGTH: 288
TYPE: PRF
ORGANISM: Bacillus anthracis
US-09-273-839A-8
Query Match 33.7%; Score 1275; DB 4; Length 288;
Best Local Similarity 100.0%; Pred. No. 1.3e-82;
Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
486 ETVARITFGKDLNIVERRIAAVNSPDLPTTKPDMTKFAKIAFGNEPNNQLOYOK 545

29 ETVARITFGKDLNIVERRIAAVNSPDLPTTKPDMTKFAKIAFGNEPNNQLOYOK 88
546 DITEPDPNDOOTSNIKOLAELNATNTYVLDKIKLNKNNILRDKRFHYDRNINAV 605
89 DITEPDPNDOOTSNIKOLAELNATNTYVLDKIKLNKNNILRDKRFHYDRNINAV 148
606 GADESUYKAEHREYINSSTEGILLNIDKIRKILSGYVEIEDEBLKRYI 665
149 GADESUYKAEHREYINSSTEGILLNIDKIRKILSGYVEIEDEBLKRYI 208
666 SSLRODKTFIDEKYNKDLPLYSNPNRYNVAATKENTINSENGDTSTNGIKIL 725
209 SSLRODKTFIDEKYNKDLPLYSNPNRYNVAATKENTINSENGDTSTNGIKIL 268
726 IFSKGYEIG 735
269 IFSKGYEIG 278
RESULT 10
US-08-960-780-32
Sequence 32, Application US/08960780
Patent No. 620435
GENERAL INFORMATION:
APPLICANT: Feltelson, Jerald S.
APPLICANT: Schepf, H. Ernest
APPLICANT: Narva, Kenneth E.
APPLICANT: Stockhoff, Brian A.
APPLICANT: Schmeits, James
APPLICANT: Loewer, David
APPLICANT: Dullum, Charles Joseph
APPLICANT: Muller-Cohn, Judy
APPLICANT: Stamp, Lisa
TITLE OF INVENTION: No. 620435: Pesticidal Toxins and Nucleotide
NUMBER OF SEQUENCES: 134
CORRESPONDENCE ADDRESS:
ADDRESS: Saliwanchik, Lloyd & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: US
ZIP: 32606-6669
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/960,780
FILING DATE: 30-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/029,848
FILING DATE: 30-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: MA-708
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 881 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: peptide
ORIGINAL SOURCE: 177c8
INDIVIDUAL ISOLATE: 177c8
US-08-960-780-32


```

QY 283 DSEFTISKNTSRRTSEVHNAEYHASFIDGSGVSAFNSNS--TVAIDHLS 339
DB 321 ----NSVSHSTNMST-----NTE-----GASVEADIGPKISGVSVNQHSET 363
QY 340 LAGERTAEFMG----LNTADTARLANIRYVNTGAPIVLPTTSVLVGNQTLATIK 395
DB 364 VAQE--WGTSTGNTSOFNTASAGTLANNVRYNNGVGAIVYKRTTISFVL--NNDIATIT 420
QY 396 AKENOLSQLAPNNYPSKMLAPIALNADDPSPPTIMANNQCLELEKTKOLRLDIDOV 455
DB 421 AKNSYALNISPESYKKGNGIAITSMDFFNSHPTILNKQYDNLNNKPMLETNOT 480
QY 456 YGNIAFYENGVRVVDGSMSEVLPJOIQTARLIFNGKDLNVERRIAANVSPDLE 515
DB 481 DG---YKIKDTHGNIVTGEMNGVIOQIKAKTASIVDGE--VAAEKRAAKDEYNEED 536
QY 516 TTKPDMLKEALKIAF--GFNEPNCNLOYGKDTPEF--NPDQTSQNIKNQIAEL-- 569
DB 537 KT-PSLTKLAKLISYDEIKTEGLIYKKNKPIYESSVMTYIDENAKKAVTKOINDTGT 595
QY 570 ---NATNIYVLDKIKLAKNNILIRDKRFRHYRNNAIAGADESVVEAKHREYINSTE 626
DB 596 KFDVSHLYV---KLTPKMWYTIK--LSILYD--AESNDNSIGKMTNTNIVSGNNG 647
QY 627 -----LLLNID-----KDIRKILSGYVEIDTE-----GLKE 654
DB 648 KKOYSSNPDANLINTDAOEKLKNRDRYISILYKSEKNTQCEITIDGRIYPTTKTVN 707
QY 655 VINDRDMINT--SSLRODKTFIDFKKYNDKLPYISNPYKVVYAATKENTIIINSE 712
DB 708 VMDONKRLDIIAHNIKSPISISSIHKT--NDEITLWLDIISI--VDVASIKPEN--LTDSE 763
QY 713 NGDF-STNGIC--KILIFSKNGYEIGZ 736
DB 764 IKOYISRYGKIKLEDOILLDKKGIIHYGE 791

RESULT 12
; Sequence 5, Application US/08471033
; Patent No. 5770696
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kozlowski, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalin M
; APPLICANT: Kostichka, N. Kistey
; APPLICANT: Duck, Nicholas B
; TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,033
; FILING DATE: 09-SEP-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: P-40,403
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 884 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-471-033-5

Query Match 20.9%, Score 790, DB 1: Length 884;
Best Local Similarity 30.4%, Pred. No. 1, He-47;
Matches 248; Conservative 130; Mismatches 292; Indels 146; Gaps 34;

QY 4 QENRLNSESSESGILGTYFEDLNFQAPVYVSTGDDLSIPSELEN--IPSENGYQ 61
DB 42 QKNO--QKENDRGILGTYFKGDF--SILTMFAPYRSTLIYKQYJANKLIDKQOEYQ 97
QY 62 SAWSGFTKVKKSDPTPATSDNHYVMYDDQEVNKNASNSKYPREGRLOYKIOYO 121
DB 98 SIMIDLOSKETDPTFNLSDDEBAILIENGKIISNKKREKQVHLEKGLVPIKIEYO 157
QY 122 RENTPTKID-----FKLYWDSNKKREVISSDNOLPELKOKS-----N 162
DB 158 SD--TFKNIDSKTEKELFKIDSGNQOQVQDELPNPEFNKESQEFIAKSKINLT 215
QY 163 SRKRSAGTPVDRNDGIDPSLEVEGTYVQKRRTPSPWISNHHKKGRYKSS 222
DB 216 QKRRKIDDD--IDTGDSDIDLWEENGTYI--QNRIVAKWDDSL--ASKGYTRFVSN 267
QY 223 PERKSPADPSYDFEKTGTGRIDKNVSPARHPVLAAYPIYVHNDENITLSKNDOSTONT 282
DB 268 PLSHVGPPTDYDEKAROLDLSNAKETFNPVLAAPSVNVSMEKYILSPENLS-- 323
QY 283 DSEFTISKNTSRHTSEVHNAEYHASFIDGSGVSAFNSNS--TVAIDHLS 339
DB 324 ----NSVSHSTNMST-----NTE-----GASVEAGIGPKISFVSVNQHSET 366
QY 340 LAGERTAEFMG----LNTADTARLANIRYVNTGAPIVLPTTSVLVGNQTLATIK 395
DB 367 VAQE--WGTSTGNTSOFNTASAGTLANNVRYNNGVGAIVYKRTTISFVL--NNDIATIT 423
QY 396 AKENOLSQLAPNNYPSKMLAPIALNADDPSPPTIMANNQCLELEKTKOLRLDIDOV 455
DB 424 AKNSYALNISPESYKKGNGIAITSMDFFNSHPTILNKQYDNLNNKPMLETNOT 483
QY 456 YGNIAFYENGVRVVDGSMSEVLPJOIQTARLIFNGKDLNVERRIAANVSPDLE 515
DB 484 DG---YKIKDTHGNIVTGEMNGVIOQIKAKTASIVDGE--VAAEKRAAKDEYNEED 536
QY 516 TTKPDMLKEALKIAF--GFNEPNCNLOYGKDTPEF--NPDQTSQNIKNQIAEL-- 569
DB 540 KT-PSLTKLAKLISYDEIKTEGLIYKKNKPIYESSVMTYIDENAKKAVTKOINDTGT 598
QY 570 ---NATNIYVLDKIKLAKNNILIRDKRFRHYRNNAIAGADESVVEAKHREYINSTE 626
DB 599 KFDVSHLYV---KLTPKMWYTIK--LSILYD--AESNDNSIGKMTNTNIVSGNNG 650
QY 627 -----LLLNID-----KDIRKILSGYVEIDTE-----GLKE 654
DB 651 KKOYSSNPDANLINTDAOEKLKNRDRYISILYKSEKNTQCEITIDGRIYPTTKTVN 710

```

QY 655 VINDRYMLN-----ISSLRDQKTFIDFKKYNKDLPLYSNPNKYVAVYATKE 704
 DB 711 VINDRYMLN-----ISSLRDQKTFIDFKKYNKDLPLYSNPNKYVAVYATKE 760
 QY 705 NTIINPSENGDT-STNGIK--KILIFSCKGYEIGZ 736
 DB 761 N--LTDSEIKOYISRGYIKLEGGILIDKKGIGHYGE 794

RESULT 13

US-08-471-044-5
 ; Sequence 5, Application US/08471044
 ; Patent No. 5840868

GENERAL INFORMATION:

APPLICANT: Warren, Gregory M
 APPLICANT: Kozziel, Michael G
 APPLICANT: Mullins, Martha A
 APPLICANT: Nye, Gordon J
 APPLICANT: Carr, Brian
 APPLICANT: Desai, Nalin M
 APPLICANT: Kostichka, N. Kristy
 APPLICANT: Duck, Nicholas B
 TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
 NUMBER OF SEQUENCES: 50
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CIBA-GEIGY Corporation
 STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: NY
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30B
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/471.044
 FILING DATE: 06-JUN-1995
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/463,483
 FILING DATE: 05-JUN-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/314,594
 FILING DATE: 09-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/218,018
 FILING DATE: 23-MAR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/037,057
 FILING DATE: 25-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Pace, Gary M.
 REGISTRATION NUMBER: 40,403
 REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8582
 TELEFAX: 919-541-8689
 INFORMATION FOR SRO ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 884 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-471-044-5

Query Match 20.9%; Score 790; DB 2; Length 884;
 Best Local Similarity 30.4%; Pred. No. 1,le-47;
 Matches 248; Conservative 130; Mismatches 292; Indels 146; Gaps 34;

QY 4 OENRLNSESSESGILGYESDLNFOADPVVTSSTGDLISIPSELEN--IPSENGYQ 61
 DB 42 QKNO--QKMDKGLGLEYFKGDF--SMITFAFTSTLIYDQVTANKLIDKQOEO 97
 QY 62 SAWSGFLTKYKSDERTFATSDNHYTMVWDDOEVKNASNSKILBLKGLYOKIOYO 121
 DB 98 SIWVIGLQSKETGFTFNLSDEQALITEINKIISNKGKQVHLKGLVYIKIEYO 157
 QY 122 RNPETKGLD-----FLYWTDSOKKREVISDMLQPLKOKSS-----N 162
 DB 158 SD--TKFNIDSKTFEKLKFKISQNGPQVQODELNPEFNKKSQELAKPKINLET 215
 QY 163 SKKRSSTAGPTVPDRDNGIPDSLEBGTVDYKNTKFLSPJLSNHEKKGLTYKTS 222
 DB 216 QMKREID--TDTGDSIPDLMEGTYT---QRIAYWMDSL-ASKGYTKYVN 267
 QY 223 PKKSTASDYSDFEKVTGRIDKNNVSPKARHPLVATFVVDHENILISNREGOSTQWT 282
 DB 268 PLESHVGDYDYDEKARDLDSNAKKEFNPLVAFPSVNVSMERKVLISNENLS---- 323
 QY 283 DETRTISKNTSTSRHTSEVHGNAEYHASFPIGGSVSAFSSNSG--TVAIDHSLS 339
 DB 324 ----NSVEHSSSTNMSYT-----NTE-----GASVEAGIGVKSISGVSVYOHSET 366
 QY 340 LAGERTAETMG-----LMTADTALNANIRYVNTGAPITVNVLETTSLVLCNQTATAT 395
 DB 367 VAQE--WGTSTGNTSQFNTASAGYLANVRYNNGTGAIDYKFTTSPVL--NNDTIATIT 423
 QY 396 AKENQSLQIAPNNYPSKNTLAPIALNADDPSTPTMYNOFLEKTKQLBLDPOV 455
 DB 424 AKSNSTALNSPESYPRKGGNGIATISMDENSHPTITLKKQVNLNKNPMLLETNGT 483
 QY 456 YENIATYFENGSRVYDTSQWMSSEVLPOIETTATIFNGKDLNVERIAVAPSPLE 515
 DB 484 DG---YKIDTHGNITVGEMNVOIQIKATASIIYDGE--RAEKRYAAKYEHPED 539
 QY 516 TKPMTLTKAKIAF--GFERNGLYOQGDTEFDE--NFDOOTSONIKNGLAEL-- 569
 DB 540 KT-PELTKALMLSTPDEIKELGELYKNNKPTLESSVMTYIDENYATREYTKLNDTIG 598
 QY 570 ---NATNIVYLDKIKLAKKANILLIDKRFHYDRNNINAVGADSVYKAEHVEVINSSTG 626
 DB 599 KFDVSHLYV---KLTPKNNVTK--LSLTDN---AESDMSIGKMTNINISGGNG 650
 QY 627 -----LLNID-----KDIRKILSGYVELEDE-----GLKE 654
 DB 651 KKQYSSNNPDANLTNTDAQKLNKNDYISLYMKSEKNTOCETIYDGEIYPTTKYVN 710
 QY 655 VINDRYMLN-----ISSLRDQKTFIDFKKYNKDLPLYSNPNKYVAVYATKE 704
 DB 711 VKDNYRDLIDIAHNKSNPISSLH-----IKTNDETLTFWMDISI--TDVASIKPE 760
 QY 705 NTIINPSENGDT-STNGIK--KILIFSCKGYEIGZ 736
 DB 761 N--LTDSEIKOYISRGYIKLEGGILIDKKGIGHYGE 794

RESULT 14

US-08-463-483A-5
 ; Sequence 5, Application US/08463483A
 ; Patent No. 5840870

GENERAL INFORMATION:

APPLICANT: Warren, Gregory M
 APPLICANT: Kozziel, Michael G
 APPLICANT: Mullins, Martha A
 APPLICANT: Nye, Gordon J
 APPLICANT: Carr, Brian
 APPLICANT: Desai, Nalin M
 APPLICANT: Kostichka, N. Kristy
 APPLICANT: Duck, Nicholas B
 TITLE OF INVENTION: No. 5849870el Pesticidal Proteins and Strains
 NUMBER OF SEQUENCES: 50

Fri, Oct 4 14:31:03 2002

us-09-848-909-21.ra

Page 11

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,483A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08 /314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA: US 08/037,057
APPLICATION NUMBER: 25-MAR-1993
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Spirull, W. Multizy
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: OCC 1695/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8613
TELEFAX: 919-541-8683
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-463-483A-5

Query Match      20.9%; Score 790; DB 2; Length 884;
Best Local Similarity 30.4%; Pred. No. 1,1e+47;
Matches 248; Conservative 130; Mismatches 292; Indels 146; Gaps 34.

4 QENRLNESSESSGGLLGYYPSDLNFOAPRVYTSSITGDLSIFSSBELN--ISENQRYQ 61
   :::::|::::::::::|::::::::::|::::::::::|::::::::::|::::::::::|
42 QKNQ--QKENDRRGLGYATFYFGXDF-SNLMPAFPRDSTLITDDQTPANKRLDKQGEVQ 97
   :::::|::::::::::|::::::::::|::::::::::|::::::::::|::::::::::|
62 SAWSGFIRYKKSDDEYTFATSNADNHVTWVDQDEVYNKASNSKRILEKGLGYIKIOYO 121
   :::::|::::::::::|::::::::::|::::::::::|::::::::::|::::::::::|
98 SIRWIGLIGSKETGDPFTFNISEBOALIEINGLIISNGKEKROYHLEKGLVIKRIEVO 157
   :::::|::::::::::|::::::::::|::::::::::|::::::::::|::::::::::|
122 RENPTEKGID-----FRLYWTDSQNKKVEYISSDNILQLPELKQKS-----N 162
   :::::|::::::::::|::::::::::|::::::::::|::::::::::|::::::::::|
158 SP-TKFNDISTFEKLEKLKRIDSONOQOVODELRNEERNKESOEPLAKPKINKMLPT 215
   :::::|::::::::::|::::::::::|::::::::::|::::::::::|::::::::::|
163 SKRKRSAGPYVPDRONDGIPDSLEVBYGVTVDKNRKRTFLSPWISNIHEKKGLTKAKS 222
   :::::|::::::::::|::::::::::|::::::::::|::::::::::|::::::::::|
216 QMKRKEIDED---PTDGDSIPLMEENGYYI---QNRIAWKWDOL-AASGTATKFSN 267
   :::::|::::::::::|::::::::::|::::::::::|::::::::::|::::::::::|
223 PEKSTASTDYSPEKYVGRIIDWNVSPEARRPIVAAYPIIVAHVMENIIISKNEOSTQNT 282
   :::::|::::::::::|::::::::::|::::::::::|::::::::::|::::::::::|
268 PLESHTYGDDYTDYEKAAROLDLSNAKETPNPLVAAFPVSVMSEKIYLSPRENLS---- 323
   :::::|::::::::::|::::::::::|::::::::::|::::::::::|::::::::::|
283 DSETTITTSKTSRSTRHSSEVHGNAEVHASFFDIGSVAGCSNMSNS---TVAIDSLIS 339
   :::::|::::::::::|::::::::::|::::::::::|::::::::::|::::::::::|
324 ----NSVSHSSTNNMSYI-----NTE-----GASVEAGIGPGKGISFEVSYNVQHSET 366
   :::::|::::::::::|::::::::::|::::::::::|::::::::::|::::::::::|
340 LAGEPTAEIWMG-----INTADTAIRANANRVRVWGTPAIYVNLPTSLVLAVGNKQTIANIK 395
   :::::|::::::::::|::::::::::|::::::::::|::::::::::|::::::::::|
367 VAOE-WGELTGITSOPTNFAAGYGLIANANRVRVNWYGCATYIVDKPTSPFLV-NNDTIATIT 423
   :::::|::::::::::|::::::::::|::::::::::|::::::::::|::::::::::|

```

[illegible]

RESULT 15
 US-08-471-046A-5
 Sequence 5, Application US/08471046A
 Patent No. 5866326
 GENERAL INFORMATION:
 APPLICANT: Warren, Gregory W
 APPLICANT: Kozziel, Michael A
 APPLICANT: Mullins, Martha G
 APPLICANT: Nye, Gordon J
 APPLICANT: Carr, Brian
 APPLICANT: Desai, Nalin M
 APPLICANT: Kostichka, N. Kristy
 APPLICANT: Duck, Nicholas B
 APPLICANT: Estruch, Juan J
 TITLE OF INVENTION: Method For Isolating Vegetative Insecticidal
 NUMBER OF SEQUENCES: 50
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 5866326artis Corporation
 STREET: 3054 Cornwallis Road
 CITY: Research Triangle Park
 STATE: NC
 COUNTRY: USA
 ZIP: 27709
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.308
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/471,046A
 FILING DATE: 06-JUN-1995
 CLASSIFICATION: 435
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/463,483
 FILING DATE: 05-JUN-1995
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/314,594
 FILING DATE: 09-SEP-1994
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/218,018
 FILING DATE: 23-MAR-1994
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/037,057

Fri Oct 4 14:31:03 2002

us-09-848-909-21.ra1

FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIY8 - SOLV4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 884 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-471-046A-5

Query Match 20.9%; Score 790; DB 2; Length 884;
Best Local Similarity 30.4%; Pred. No. 1,1e-47;
Matches 248; Conservative 130; Mismatches 292; Indels 146; Gaps 34;

OY 4 QENRLNLESSSOGILGYVPSDLNFOAMVYTSSTDLSIPSELEN--IPSENOYFO 61
DB 42 QKNQ---QKENDKGLGYFKGDF-SNLTMPAPTRBSTLIYDQOTANKLIDKQOEYO 97
OY 62 SAIVSGFIRKVKDEYTPATSDNHVYMWVDOEVINKASNSKIRLEKGRLOYKIOYO 121
DB 98 SIRWIGLQSKKEGDFPLMSDEDAIIEINGKIISNCKEKQVHLEKGLVPIKIEYO 157
OY 122 RENPTEKGLD-----FKIYTDSONKKEVISSDNLOLPELKOKS-----N 162
DB 158 SD--TKFNIDSKTFKELKFKIDSONOPQOVQDELBNPFNKKESEFLAKPSKINLFT 215
OY 163 SRKKSSTAGPTVDRDNGCIPDSLEVEGYTVYVKNKTFPLSPWISNIEHKKGLTKXSS 222
DB 216 QKKKREIDED---TDYDGSIPDLMEENGIT---QNRIVAKMDSL-ASKGYTKFVSN 267
OY 223 PEKMWSTADPYSDPEKVTGRIDKNSPAPRHPLVAAIPIVHVMENIISKNEOSTONT 282
DB 268 PLSHTVSGPTDYDEKAAARDLISNAKETFPVLAAPSPVANSMEKYLSPNENS---- 323
OY 283 DSEFTISKNTSTRTYTSVHGNAEVHASFDDIGSVSAGFSNSNS---TVAIDHLS 339
DB 324 ---NSVESHSSTWMTY---NTE-----GASVPAIGPKGISPGSVNQHSET 366
OY 340 IACERTMAETMG---LNTADTARLANIRVNTGTAPIYVNLPTSLVLGKQTLATIK 395
DB 367 VAQE--WGTSCTNSQPFNTASAGYINAVRYNNNGGAIYDKPTSPVL--NRDTIATIT 423
OY 396 AKENOUSOLIANNYYPSKMLAPIALNADDSSTPTIMYNOELEKTKQLRLDTDOV 455
DB 424 AKSNSTALNISGESYPRKGGONGIATTSMDPNSHPYTLNKKQVNDLNNKPMLETNOT 483
OY 456 YGNLATYFNENGVRVVDTSNMSEVLPJOETARLIIFNGKDLNVERRIAANVPSDLE 515
DB 484 DG---YKIKDGHGNIYVGGMGNGYIOIKAKTASIIYVDGE--RYAEKRYAARDYENPED 539
OY 516 TTKEDMTLKAALKIAF--GENEPNGLOYOGDTEEDF--NPDQTSQNIKNQLAEL-- 569
DB 540 KT-PSILKNAKLISYDEIKETLEGLLYKKNKPIYESVATYIDENYAKETKQLNDTGG 598
OY 570 ---NATNIYVADKTKLAKKNLILDRKFRHVDNRNIAVGADESVYKAEHREVINSSTEG 626
DB 599 KKKDVSHTIYV---KLPKKNVYTIK-LSIILDN---AESNDSIGKWTNNTIYSGGNG 650
OY 627 -----LLMID-----KDIRKILSGYVLEIDTE-----GLAKE 654
DB 651 KKQYSSNPNPAULILMTDQEKLNKNDYISLYMKSEKNTQCEITIDGELIYPTTKTVN 710
OY 655 VINDRTDMLN-----TSSLRDQKFTIDFKKYNDKPLPYISNPNYVAVYATKE 704
DB 711 VNKDNYKRLDILAHNIKSNPISLH-----IKTNDLITLPMDDIST--TVASIKPE 760

OY 705 NTLINSENGDT-STNGIK---KIIPSKKYEIEIG 736
DB 761 N--LNDSEIKOYISRGIKLEDEGILLIDKKGSLHGE 794

Search completed: October 4, 2002, 14:34:45
Job time: 30 sec

Fri Oct 4 14:31:03 2002

us-09-848-909-21.ra1

```

OY 519 PMWLKALRIAFGNP-----NGNLOYGRDITFDFNFDOOTSNIKNOALNATNIT 575
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 932 ISDIRNSLKTIIOFSESIYNDIKELKEKNVLE-----SONNNTDINOYLSKIE--NIT 984
OY 576 TVLDKIKLNAMNILLIDKRF--HYDRNNIAGADESVKAEHREVINSTEGLLNDIK 633
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 985 NIL-----KLNKIKKIIDKVEYDELEKNN-----KKNALNSN----- 1020
OY 634 DIRKIIAGYIEIDTEGLKE-----VINDRY--DMLNSSLRODKTEFIDKRYND 663
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1021 --EKI-----ITOLKENSLEKECOSKIKSTIDNYSCEIKNTYL-----KTYIVAKEN-- 1069
OY 684 KLPITYSN--PNKRYNV-----YAVTKENTYINSENGDTSTN--GIKKILIFSCK 730
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1070 -IMTYFRNAEYNONVSLNFNNEIMADYSOYLINIKNNKNTNMDYINIKELKEHKK 1126

```

RESULT 8

B64635 toxin-like outer membrane protein HP0922 - Helicobacter pylori (strain 26695)

C:Species: Helicobacter pylori

C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999

C:Accession: B64635

R:Tombl, J.F.; White, O.; Kervatage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodok, A.; McKend

son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watney, L.

Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Fraser, C.

A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A:Reference number: A64520; M01D:97394467

A:Accession: B64635

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2529 <TOM>

A:Cross-references: GB:A6000602; GB:A6000511; M1D:g2314060; P1D:AAD07969.1; P1D:g231406

Query Match 4.98; Score 183.5; DB 2; Length 2529;

Best Local Similarity 21.08; Pred. No. 0.47; Mismatches 261; Indels 241; Gaps 36;

Matches 156; Conservative 94; Mismatches 261; Indels 241; Gaps 36;

```

OY 123 EMPTEKGLDRIKYWDSONK-----KEYISDMLOPELKSSNRKRSKAGPVPDR 178
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 729 QNRPASVMSGYTL--QNKTYSMKGIYID-----PNLSQSGSGSGLTSTYFANLF-- 779
OY 179 DNDGIPDSLEEGYTDVKNKRFSLP-----WT-----SNHKKGLTKYK 220
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 780 ---GSPFVNIOGTLIIIGNTESVNSGLWIGGFGYITGFSAANIT---LNMF 832
OY 221 SSPEKMT-----ASD-----PYSPEKVTGRIDKNVSPKARHPVAAPIY 262
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 833 KIDGVSNSDGGANITFKASDNITMDGLNYDAETVTKMIOGAS---QHSVATFDALN 889
OY 263 HYDMNITLISNEPOSTONTDEPRTISKNTSRTHTSEVHNAEYHASFIDIGSVSA 322
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 690 NISVNSSES-----DMTWCKFEFSKKNISFS-----NSF-----S 921
OY 323 GFNSNSSTVA---IDHSLGERTVAETMGANTADARLANIRYVNTGAPITVNL- 378
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 922 GFNPGSGSVISANATNSLFSNR-----LNGAVYNIQANSLTFN--NQOATFVNL 973
OY 379 -----PTTSVIAGKNOTLATIKAKENOLSOILAPNNYPSKNIAPLAL--NMDDFS 428
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 974 SROGSFNATITOLLGNTNFTLS-----QSLNFMNDTTLNANNITLANKSOAFK 1025
OY 429 STPTIYNYOFOLEKTKOLRLDQOYGNATYNNENGRRVAVDTGSMSEVLPOLIOET 488
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1026 NS-LTIDNNSINSLDNOGYVNNANNTSANNOSASLNTNGS-----QATF 1068
OY 489 ARIIFGKDLNL--VERRIAAVNPSPLETTYKPMWTKALRIAFGNPKNLOYGKDI 547
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1069 NSLTFNGTSLNMSKLNNSNFSNMT---INLDDSVLSASNTSLANINFGASQ 1125

```

```

OY 548 TEF-----DENEDOOTSONIKNOA-----E 568
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1126 ADFGNTITIDTASFENDDASSLNPNNTLFGALNENGTYPSL/KALMSVSGFVLGNNG 1185
OY 569 LNATNTYVDKRIKLNAMNILLIDKRF--HYDRNNIAGADESVKAEHREVINSTEGL 605
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1186 INLSOI--NIFDNITKSVTYNILLNMOKIGITGSGANGCYERILFYKKNIONATYSDNNINOT 1244
OY 606 GA-----DESUYKAEHR-----EVIN--STEGILLNIDKDI-----RKILSGYVE 645
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1245 WSPINLMSQIIOESIKNGDLTIEVLNPNNSASNTIFNIAEYNYOASKONPGSYD 1304
OY 646 IEDTEGLKEVINDRYDMLNSSL-----RODKTEFIDKRYNDKLPY--ISNENY 694
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1305 YSDNQA-----GTYLTSNKKGLFTPKGOSOTPOAPQYSPFNQTLSSLNINPKGFSBNL 1359
OY 695 KANVYAVKENTYIN--PSNGDPTSTNGIKKIL 725
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1360 KTLIGLISONSATLKEMIESNOLDNITN--INEVL 1392

```

RESULT 9

T31102 filamentous hemagglutinin 1 - Haemophilus ducreyi

C:Species: Haemophilus ducreyi

C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999

C:Accession: T31102

R:Ward, C.K.; Lumley, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.

J. Bacteriol. 180, 6013-6022, 1998

A:Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.

A:Reference number: 220984; M01D:99030326

A:Accession: T31102

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-4152 <HMR>

A:Cross-references: EMBL:AF057695; M1D:g3929017; P1D:g3929018; P1D:MAC79757.1

C:Genetics:

A:Gene: lspA1

Query Match 4.88; Score 183; DB 2; Length 4152;

Best Local Similarity 22.28; Pred. No. 0; Mismatches 321; Indels 218; Gaps 45;

Matches 187; Conservative 118; Mismatches 321; Indels 218; Gaps 45;

```

OY 10 NESSESGILGYTSDLRQ--APMYTSTGD-----LSIPSELSNIPSENOYE 60
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 162 SKTESISQ--LGLKLANIQLQKEAKKLINQVGDHESNIOGALEVAGKRAADLIIV-- 216
OY 61 GSATMSGRIVAKKSDYTPATSADNHTVWVDQGVINKASN-----SNKIRLEKGRLY 114
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 217 PGLITLANKYITINDRFVYSTS-----DIIHRENGILSVRNGKVTIDKGVVA 264
OY 115 QIKIOY---QRENPTKGLDPLKYWDSONKKEVLSDD-----NLQPLKOKSS 161
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 265 TNGLSHFPEYVAHNDKGG---KIVAKTENOKSVNPNANITPAAGSLNMYLKTREATPIS 321
OY 162 NSKKRRTSGAPVPPRDNDGIPDSLEVEGYTD-----VKNKRFELSPMISNIEHKKGLT 217
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 322 GTR---TSDPPI--SADAGSMGVSNIKFEVYDKAGVKKHGIIFSEBNDIKKWDGNA 377
OY 218 KYSSPEKMGWASAPYSDEPKYTGRIKNSVPEARHPVLAAYPIVHMDMENTILISKMDQ 277
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 378 SLKEIYAK-----KDIDLAKDIELEKCOLANNNKILILSTGKINLNKASEV 425
OY 278 STONDSERTIS--KNTSTSRHTSEVHNAEYHASFIDIGSVSAGFSNNSSTVALID 336
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 426 SADVAVASEMLALENSMSANSLDVITKIEVNRS-----SKVSAGTATKASNTITD 480
OY 337 SLIAGEPRTV-----AETMGINTADPARLN-----ANIRYVN---T 369
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 481 SSVAVANKITLVNTNNTANNOSKLSAKMDIELVTHNITLNTS--KLSAQKANKITENTLTN 540

```


Best Local Similarity 18.7%; Pred. No. 0.54; Indels 214; Gaps 36;
Matches 158; Conservative 140; Mismatches 311;

```

QY 1 ENKQRL-----LNESSSSGLGYYFSDLNFOAPMVVTSSTGDLSPSELENIPSE 56
DB 485 QYLNKLTFFESLNHNHANNNELI-KYPSDLKANLQINEENKLNQTEKTEKFNDEIK 543
QY 57 NQYFQALMSGFIYKKSDEYFATSANHYVMWD-----OEVIKASNSNRIKLEK 110
DB 544 NIHNIEIKETIKIHS-IYNSIEFEIREIGINSLTAKFEKVENYTNINIKR-EK 601
QY 111 GRLYQIKTOYQRENTEKGLDFYLYTQSNK-----KEVSSDNLQ----- 152
DB 602 LKHTDF-----SDPKGNIKYTKTKIKKINDIMAVSQDIOHINQIDIDIOKSES 652
QY 153 -LPELKOKSSNRKRSKTSAGPTVPDRNDGIPDSLEVEGYVYVKKRFFELSPMISNI 211
DB 653 YVSEMKQIINKLEKVSNT-----EISND-----NVEG-----LKKQOIT---VTKID 692
QY 212 ERKG-----LTKYKSPKSTASDPYSDPEKVTG-----RDKVNSPEARHPL 255
DB 693 KKKNIYEIKRLISEIKEDNTSEKVKYDINLSYQNLGNLFLEQIDEE-KKKAKNTI 751
QY 256 VAAAPYIVHVMENILSKNEQOSTONTDETRTISKNTSRTHTSEVHGAEVH---A 311
DB 752 KMEWAYID-DLDMIRKKSQELIETEMDKMDINKEMALKISHDDOKCHDSKHKHENS 810
QY 312 SFEDIGSVAGFSNSNSTVAIDSLAGERTWAEGLNTADTAALNANIRYVNTG 371
DB 811 DLYDKSSKTIIDFSR-ESDINDIKNLQ-----KNVSESQNHNDIQCINEY 857
QY 372 APIYVWLPYTSIVGKNOTLATIRAKENOLSOILAPNNYPSK-----NLAP 418
DB 858 ANIYML-----KLNKIKKIDIKVEYTSIELEKKNKINDELANSER 899
QY 419 IALNMODESTPTIYVNOFLE-----LEKTKOLRLDQ-----QYGN 458
DB 900 VIKKIEGSLSKECRSKINSTLDKIDECIKININLKNLNEINININHEKNAEELK 959
QY 459 IATYNEGRVAVDQSNSEVLPJOIETARIIIFNGKDLVERRIAVANPSDPLETK 518
DB 960 IYLSNKN-----IMADNKSQYILIKKNGT---NDHYNIKELK-SHKDSNCKY-TE 1010
QY 519 PDWTLKALKIAPFNPNNGNLOYGRD-----ITFEFDPFOQTSQNIKQALAE 568
DB 1011 ADONKAIOR-----NKELFQYKEEVYVLNKYVAVELKNKFP-KTKNDSKOIIRE 1061
QY 569 IMAITTYVLDKIKILAKKN-----ILIRDRFYDRNNIIVAGADESVYKEAREVIN- 621
DB 1062 IKDHNICTLESCKSEKKEKNEIKNEKIHEDEVANDKSKAKITSIKYSEPEPKTIKI 1121
QY 622 ---STEGLLNIDKIRKILSGYVEIEDT-----EGKEVINDRYDMLNLSIR 669
DB 1122 NEIKTSDDCKEYNDLEKQISLSDIOETKLENGKQKLTLEL-----LESIK 1172
QY 670 QDGTETIDEKTYNDKLPYISN-----PNKVVAVYVRENTIIPSENGDSTNG 720
DB 1173 KKKNIIDOKREIDVNSKINIENTVNOHKNKNEYIGI---VEKINELAKTKNQISSTRE 1230
QY 721 IKK 723
DB 1231 LIK 1233

```

RESULT 12
S05603
major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (st
N:Alternate names: gp195 surface antigen
C:Species: Plasmodium falciparum
C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jun-2000
C:Accession: S05603; S04850
R:Myler, P.J.
Submitted to the EMBL Data Library, April 1989

A:Reference number: S05603
A:Accession: S05603
A:Molecule type: mRNA
A:Residues: 1-1639 <MYL>
A:Cross-references: EMBL:X15063; NID:g9896; PDB:CA33163.1; PID:g9897
R:Myler, P.J.
Nucleic Acids Res. 17, 5401, 1989
A:Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from
A:Reference number: S04850; MUID:89345116
A:Accession: S04850
A:Molecule type: mRNA
A:Residues: 1504-1639 <MYL2>
A:Cross-references: EMBL:X15063
C:Superfamily: major merozoite surface antigen
C:Keywords: glycoprotein; merozoite; surface antigen
F:1.19/Domain: signal sequence #status predicted <Sig>
F:20-1639/Product: major merozoite surface antigen #status predicted <Mat>

Query Match 4.8%; Score 179.5; DB 2; Length 1639;

Best Local Similarity 20.4%; Pred. No. 0.39; Indels 285; Gaps 41;
Matches 171; Conservative 111; Mismatches 272;

```

QY 5 ENRL-LNESSSSGLGYYFSDLNFOAPMVVTSSTGDLSPSELENIPSENOYFOSA 63
DB 927 ENILSLGKKNKIYQELIGKSSK-NF-----YKLIKDSDFYNE 965
QY 64 IWSGFIYKKSDEYFATSADNHVMWDQOEVIKASNS-----NKIRLEKGRY- 114
DB 966 SETFVFKSKAD-----INSLNDSKRRKLTEDINKTKTLQISFDLYNKTKIKLERLD 1020
QY 115 -----QIKYOARENTEKGLDPKLYMTDSQNKKEVYSSDNLOLPELKOKSSNSRKRK 168
DB 1021 KKKTVGYKKMOIKKTLTKLEDESKL--NSLNPKHYL--QNFVFNKKKEALIEATEN 1076
QY 169 TSAGPTVPDRNDGIPDSLEVEGYVYVKKRFFELSPMISNIHKKGLTKY--KSSP- 223
DB 1077 T-----LENTKILKHY-----KGLVYKNGESSPLK 1103
QY 224 ---EKWSTASDPYSDE-----KYGRIDKAVSPEAR-----HFLVAAPIYVHD 265
DB 1104 TISESIOETEDNTASLENEKYLSKLEKLDONLNEKKKLSYLSQJHLHLA-----E 1156
QY 266 MENILSKNEQOSTONTDETRTISKNTSRTHTSEVHGAEVHASFEDIGSVAGFS 325
DB 1157 LKEVIKKNK-----YGNSPSENNY-----DVNNALESIKKFLPEGTDAVAVS 1200
QY 326 NSNSSTVAIDSLAGERTWAEGLNTADTAALNANIRYVNTGPIYVNLPTT----- 381
DB 1201 EGSQDTLPQSPKRPASTVGAES---NITTSQ--NVDEYDVYIVPIIGSESDYDDL 1256
QY 382 -SLVYGKNOTLATIRAKENOLSOILAPNNYPSKNAPIALNADQDFSSPTIMANTNOL 440
DB 1257 GQYVGEAVTSYV---DNILSKI---ENEYEVLYLKPLA-----GYR 1294
QY 441 ELEKTKQLADLDQVYGNATYVFNENGRV---RVDTGSNWSSEVLPJOIETARIIIFNGK 497
DB 1295 SLRK-----QLENNVATFVAVNKDILNSRKNREKFNKVL------SD 1332
QY 498 L---NLYERRIAVANPSDPLETKRPDMLKALKIAPFNPNNGNLOYGRDITEFDN 553
DB 1333 LIPYKDISSNYYVADPYKFLNKKERDKLSYVI-----KSDITFDIN 1377
QY 554 PDQQT-----SONIKNOLAEI-----NATNLY-TVLDKI----- 581
DB 1378 FANDVLYTKILSEKTKSDSDSIKTIYNDKGENEKYLPPLNINIELKTYVNDKIDDFVI 1437
QY 582 -----KLNKAKNTI-----LIRDRFYDRNNIIVAGADESVYKEAREVINS 622
DB 1438 HLEKAVLNATYKESNVEYKIKELNLYAKTIQDGLDFKKNNNFVGIADLSDTYNNINILK 1497
QY 623 ---STEGLLNIDKDI-KRILSGYI---VEIETDEKLEVINDRYDMLNLSIRDQGTET 676

```

Db 1498 FLATGAVFENLAKTVLSNLLDGNLQMLNLSOHCYKCKKOPNCGCPRIHDERECCKL 1557
 QY 677 DEKYNDKPLPLISNPNYKAVYAVKENTJINSENG-----PSTNGIKKI 724
 Db 1558 NYKQGDGKC---VENPPTCN-----NCGCDADAKCHERDSSGSG-KKI 1599

RESULT 13
 A:Accession: A86827
 C:Species: Lactococcus lactis subsp. lactis (strain IL1403)
 C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
 R:Botulin, A.; Mincer, P.; Manger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrlich
 Genome Res. 11, 731-753, 2001
 A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s.s.
 A:Reference number: A86825; MUID:21235186; PMID:11337471
 A:Accession: A86827
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1072 <STO>
 A:Cross-references: GB:AE005176; PID:912724625; PIDN:AA05715.1; GSPDB:GN00146
 A:Experimental source: strain IL1403
 C:Genetics:
 A:Gene: y4fg

Query Match 4.7%; Score 178.5; DB 2; Length 1072;
 Best local similarity 19.4%; Pred. No. 0.24; Indels 179; Gaps 30;
 Matches 160; Conservative 142; Mismatches 342

2 VKQENRLNLSSESSOGILGYPSDLAFQAPVYTSSTIGSLSPSEELN-----52
 Db 300 VEQYDVASSSETPQDANASALYPISEASVYDNTLMSLSLSSISSGTFENSGASSTA 359
 QY 53 ---IPSNQYFOA---IMSGFTKVKDEYTFATSDNRYVMWDOEYVNNKASNKI 106
 Db 360 ELYSDSESNLSLSSNQINSNSSEKDSKOSLSSGSESEHSNSNINETNNSEI 419
 QY 107 R-----LEKRLYQIKIYQORENTEKGLDFL---YMTDSQNKKEVYSSDNOLPELK 157
 Db 420 TILPSPNTESSNSVSDQSSASTNSISLSPNSISSTQSEANSSDFENVAEVA 479
 QY 158 OKSNSRRKRTSAGPTVPDRNDGT---PSLEVEGYTVQVKNRFTPLSPWISNHEK 213
 Db 480 NNSLASVNNSSSVLSTSTADNIGINOSGDILKD---SEISTSGAPLS---SNQTS 534
 QY 214 KGLTKYKS---SPEKWS-----TASDPYDEKYTVGRIDKVNPPARPIVAAVPI 261
 Db 535 EASTNSNSISLSPNISLSTVLESTISSNPNVAEVAANNSLASVNNSSSVLSTSTA 594
 QY 262 VHYDM-----ENT-----ILSKNDQSTQNDSEFTPTISKNTSTRTSTSEVH 304
 Db 595 DNEIINQFSGDMLTKRDSSETISGAFLSNQTSSEKSSNSINSPLSLSTLSTNSA 654
 QY 305 GNAEVAHAFPDIGGSVAGSNSNSTVAIDHSLSL---AGEETAFAEMGLNADTARLNA 362
 Db 655 TNO-----SNSPEATKVDNNSST---HSSNINLNGSNDSDSDSDSDSSNL-- 699
 QY 363 NIRVNTGTAPLYNVLPSTISLVGRKQTLATIKKENQSQ-----ILAPNNYPSKNLAP 418
 Db 700 -----SSSNLETPNQIISKPEVNNISEPKKVSNSVENDTDE 742
 QY 419 IALNADDPSTPTITMANTNOLEKTKQLRLDDQYGYATYV---FENGVRVDTGS 475
 Db 743 MESTNKSISPISTISSQKESQSN---LLNTTEGINPIYFNNSSEKSSASTLT-- 797
 QY 476 NMSEVLPQIOETTAIIP-----NGKDL-----NLYERRIAVNSDPLETFKPM 522
 Db 798 SYSNSSSSSTGCLYSNEAGRDNGSEHSLPSSNSNENNVSSISQOALLESKSTN 857
 QY 523 LK-EALKAIFGFNP-NGNLQYQKDTEDPFPQOQTSNIRKNOLAEINATNYVLAK 580

Db 858 KSSSLSTINSTSHPNEDNOSNSD-----EYKSNNSVESITIGOLNLSIKNTKMS 909
 QY 581 I---KINKNMILLRDKRPHYDRNNIAVAGDESVAKEHREVINSTBELINDIDIRK 637
 Db 910 LTKOKLSIYVILPESKSVTNEKNSNVTSEELIKITPKQK-DKSONLQITALLDSFNK 968
 QY 638 ILSGYVELEDTEPELKEVIN-----DHYDMLNLSLROD-TRFIDFKYNDKPLX 688
 Db 969 -----EYETMEDSKTVPDKVLDNEDNOSNNKSTIADKMKVFPKRSEFNKTL- 1019
 QY 669 ISNPKNVAVYAVKENTJINSENGDTSTNGIKTILPSSKC 731
 Db 1020 -----DSDNMLIKTVILKKG 1036

RESULT 14
 A:Accession: G90599
 C:Species: Mycoplasma pulmonis
 C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
 R:Chandaud, I.; Helli9, R.; Ferris, S.; Barbe, V.; Samson, D.; Galissou, F.; Moszer,
 Nucleic Acids Res. 29, 2145-2153, 2001
 A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma
 A:Reference number: A99512; MUID:21267165; PMID:11353084
 A:Accession: G90599
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1752 <KUR>
 A:Cross-references: GB:AL445566; PID:914090118; PIDN:CA013876.1; GSPDB:GN00153
 A:Experimental source: strain VAB CTIP
 C:Genetics:
 A:Gene: WPU 7030
 A:Genetic code: SGC3

Query Match 4.7%; Score 178; DB 2; Length 752;
 Best local similarity 19.4%; Pred. No. 0.15; Indels 262; Gaps 41;
 Matches 160; Conservative 125; Mismatches 277

41 GDLSPSELENIPESENOYFOSAIMSGFIKYSDEYTF-----ATSDAHHTVMD 92
 Db 20 GTISGISYGISOLPNESASL-----VKRANLMLNKNSTYLNKSTIELINQMD 69
 QY 93 DOEYINKASN---SNKIRLEKGR-LYQT---KLOYORENTEKGLDFLYWDSQ 140
 Db 70 EKVYNSISANDFPNKFVQSKTPLENGEKTITVSLGRDIYFQVLYNSQKITSKSVKITSK 129
 QY 141 NKKEVYSDMLQPLKOKSSN-SRKRSTISAGPTVPDRND---GIPSLVEGYTV 194
 Db 130 ISKDVAMD-----KORLNDPANKLNVNFKSSSEKQSDYVAQSFNDKSKLEIKYL 182
 QY 195 DVKN-----KRTPLSPWISNHEK---GLTYKSPPEKMTASDPY 233
 Db 183 DKNVEKNISENTEFVELKTENNAFVGGISNDIGTKVLEATVYKKNNSK-----Y 235
 QY 234 SPEKVTGRIDKNVSEARHPLVAAYTYVDMENILISKNEQDSQJNDSEFTRTSKMT 293
 Db 236 ONIOKI---QITNNE-----KRFQISDESILKELNVAVV-- 266
 QY 294 STSRHTSEVGNAEVHAFPDIGGSVAGSNSNSTVAIDHSLSLAGEERTMAETGLN 353
 Db 267 -----TEBQKSKSEMDASFS-GNSLVSKEKSLND---DEKIKLES---YEVVDLK 312
 QY 354 TADTARLNANIRVYVNTGTAPLYNVLPSTISLVGRKQTLATIKAK-----ENLQITL 405
 Db 313 TNEWTLSYKRDHK-----FKLEIYEQKSGVATLYLLGMVAVNGESIPNRLSLF 366
 QY 406 APNNYPSKNLAPLALNADDPSTPTITMANTNOLEKTKQLRLDDQYGYATYVFE 465
 Db 367 RPS---SKESTIITIGKELSREKLINSIVKVELKMTSROAD-KILASTISKQFD 421
 QY 466 ---NGVRVDTGSN-----WSEV-LPQIOETTA---RIIFNGDLNLVERRIA 506

```

Db 422 OSQANKRITIGKRAMTDIAKLAKALYTELSVENINDTGLIKLYFLKADDSITKRTT 481
OY 507 AV-NPSPLETETKPD-----MTLEKALIAFGNENGMLQYQOKDTERDFDPQOTS 559
Db 482 LVLSGFKAVATKPEDEKMDVEIKKSLMOIT-----YDDP----- 517
OY 560 QIKKQIAELNATNIY--TVLDR-----IKIAKMLILRDKREHYDR 600
Db 518 -----LMTINLYKNTFTLSKELNKRIRQISDEDYFDIKILNTNKEKEDRDLFDS 568
OY 601 NNIAVGADESIVKAREVINSSTEGILLNIDKDIRKILSGYIEIEDEG-LKEVINDR 659
Db 569 NEKAI-----EYKNNDSSTLVNFKYIKVKK---VHSDKSSGKIEDIYVDR 613
OY 660 YDMLNISLRDQ-KYFIDP-----KRYNDLP-----LYISNPNKYVNYA 700
Db 614 VESKEVKKNIGNAAKRFQDSLSKDSQRLKPLSQITTNELKDQISVDEPKDILISSN 673
OY 701 VTEN-----TIINPSE--NGDTSTNGIKKIL-----FSK 729
Db 674 INRDNPMFEYKILIEKDKDNAD-DKNGSIKIMVSLYKNTKFSK 716

RESULT 15
Jc6009
surface-located membrane protein Imp3 precursor - Mycoplasma hominis
C:Species: Mycoplasma hominis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: Jc6009
R:Label: S.A.; Jensen, L.T.; Brock, B.; Birkelund, S.; Christensen, G.
J. Bacteriol. 178, 2775-2784, 1996
A:Title: Analysis of 0.5 kb chromosome pair repeats in the Mycoplasma hominis Imp gene system
A:Reference number: Jc6009; M01D:96213016
A:Accession: Jc6009
A:Molecule type: DNA
A:Residues: 1-1302 <LAD>
A:Cross-references: EMBL:X95601; NID:g1197335; PIDN:CA64858.1; PID:g1197336
C:Genetics:
A:Gene: Imp3
A:Genetic code: SGC3
C:Superfamily: surface-located membrane protein Imp3; tetratricopeptide repeat homology
C:Keywords: duplication; membrane protein
F:1-24/Domain: signal sequence #status predicted <SIG>
F:1-25-1302/Product: surface-located membrane protein Imp3 #status predicted <MAT>
F:957-992/Domain: tetratricopeptide repeat homology <TP1>
F:993-1026/Domain: tetratricopeptide repeat homology <TP2>
F:1089-1120/Domain: tetratricopeptide repeat homology <TP3>
F:1134-1190/Domain: tetratricopeptide repeat homology <TP4>

Query Match 4.7%; Score 178; DB 1; Length 1302;
Best local Similarity 17.9%; Pred. No. 0.33;
Matches 137; Conservative 141; Mismatches 287; Indels 200; Gaps 27;

OY 1 EYKQENLNESSSSQGLGYFDLNFQAPMVVTSSTYGDLSIPSSLENIPEENYOF 60
Db 476 QYDENKSTIEQNLMLDKANTLLPOLINDSEIYKAKESLNAETTNANKRNVNDNNSM 535
OY 61 QSAIMSGFIYKKSDEYFATSDNVHTWVDDDEVINKASNSNRIEKKR-LYQIKI 118
Db 536 QSA-----KSLDDKVTIKQNLTERFKDKDAKEKELEQFRKDIDNFLT 579
OY 119 QYQENPTEKGL-----DFKLYTDSQNKKEVSSDNILPDLKOKSSNSR--KRS 168
Db 580 DQYKRPVATLVKDLNKAQDKKSVTKSSNKEIIAND---ELKQALDKAVAKQI 635
OY 169 TSAGPVDNDNGIPDLSLEVGTYV---DVKNRRTFLSPWISNIEKKGL----- 216
Db 636 DEKANSIKQSLDSITNANQLINKLVSDKDIQAKTELISOASQELNLANPFSMOS 695
OY 217 -----TKYKSSPEKSTASD-PYSDEKVTGRIDKNVSPARPLVAHPIVAVDM 266

```

```

Db 696 AKESIDAKYTELTETKLETFENKDKDVKEKELEKTRKDIDEFINTKELNP----- 743
OY 267 ENILSKNDOSTONTDSEFRTISKNTSTKRTTSEVHGNAEVAHSFEDIGGSVAGPSN 326
Db 744 -----NYSLTISELTSKRDSKSVTNSNSKSDI----- 771
OY 327 SNSSTVAIDHSISLAG-ERTWAEFMGLMTADTARLANIRVNTGAPINVLPTTSVL 385
Db 772 -ETANTELQALAKANDKQADNLAKSTKE--QLNNSISSANLILAKLD----- 819
OY 386 GKNOTLATIKA---KENQSLQILAPNNYPSKNLAPALAO-DFSSTPTNNYQPLE 441
Db 820 -KDNITQAKTELEKEVOKANQAVASNTASQSKSLAKYTELTKELETFENKDKVK 878
OY 442 LEKTRQRLDTPDQVGNATATYNEENGRAVDGTSNWEVLPOLOETFARIITFNKDLNLY 501
Db 879 FKELEQTRKIDEE-----INTKTNPDYSTLISELTK-----R 913
OY 502 ERRIAVNPSPDPLETTPKPDMLKALKIAGFNEBNGMLQYQOKDTERDFDPQOTSQ 561
Db 914 DSKNSITNSKSDIETANTELKQALAKA-----NFDKQAD-NLARSTREQ 959
OY 562 IKKQIAELNATNIYTVLDRKIKLAKMMLILRDKREHYDRNNIAVGAD---ESVKEAARE 618
Db 960 LKNSISSAN-----TLIAKIT-----DKDNITQAKTELEKEVOKANQ 998
OY 619 VINSSTEGILLNIDKDIRKILSGYIEIEDTEGLKEVINDR-YDMNISLRDQKTFID 677
Db 999 VASNNATSM-----QSKSSLDKAVTEL--TKLETFENKDKVATFELQFRKDIDEFIN 1051
OY 678 FKYKNDKLPVLYSNPNKYVAVT---KENTIDNPSNGTSS 718
Db 1052 TNK-----TNPNSYTLISELTSKSDSKNSTNSKSDIER 1087

```

Search completed: October 4, 2002, 14:35:21
 Job time: 66 sec

Fri Oct 4 14:31:04 2002

us-09-848-909-21.rpt
